

Open Research Online

The Open University's repository of research publications
and other research outputs

Functional characterisation of a novel arginine/serine-rich splicing factor

Thesis

How to cite:

Cazalla, Demián (2003). Functional characterisation of a novel arginine/serine-rich splicing factor. PhD thesis
The Open University.

For guidance on citations see [FAQs](#).

© 2003 Demián Cazalla

Version: Version of Record

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.21954/ou.ro.0000f55d>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

**Functional characterisation of a novel
arginine/serine-rich splicing factor**

Demián Cazalla

Doctor of Philosophy

The Open University

2003

**Sponsoring establishment: MRC Human Genetics Unit,
Edinburgh**

Submission date: 1 September 2003
Award date: 14 November 2003

ProQuest Number:27527241

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 27527241

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Table of Contents

Table of Contents.....	1
Table of Figures.....	4
Declaration.....	5
Acknowledgements	6
Abstract.....	7
Abbreviations used	8
Introduction.....	11
1.1. Gene expression and RNA processing.....	11
1.2. Splicing	12
1.3. Alternative splicing.....	17
1.4. The spliceosome	20
1.5. SR proteins	25
1.6. The RNA recognition motif	30
1.7. The RS domain	32
1.8. SR protein phosphorylation.....	34
1.9. Functions of SR proteins	36
<i>Roles in splicing.....</i>	<i>36</i>
<i>Exon-dependent functions</i>	<i>36</i>
<i>Exon-independent functions</i>	<i>39</i>
<i>Other functions of SR proteins.....</i>	<i>40</i>
1.10. SR-related proteins.....	41
<i>SR-related proteins and splicing</i>	<i>41</i>
<i>SR-related proteins and the transcriptional machinery.....</i>	<i>46</i>
<i>SR-related proteins and 3'-end processing</i>	<i>47</i>
<i>SR-related proteins and cell structure</i>	<i>48</i>
1.11. Identification of new factors involved in RNA processing.....	49
1.12. Conclusions	51
Materials and Methods.....	52
2.1. Plasmid constructs	52

2.1.1. Cloning and sequencing of mUSSRp58 and hUSSRp58	52
2.1.2. Epitope-tagged expression plasmids	53
2.1.3. Plasmids used for yeast two-hybrid analysis	55
2.1.4. Plasmids for expression in Baculovirus system.....	57
2.2. Antibodies	57
2.2.1. Production and purification of antibodies against USSRp58	57
2.2.2. Other antibodies	59
2.3. Sequencing of DNA	60
2.4. Northern blot	60
2.5. Western blot.....	61
2.6. Cell culture and transfections	62
2.7. Indirect immunofluorescence	63
2.8. Nucleocytoplasmic shuttling assays	63
2.9. Yeast two-hybrid	64
2.9.1. Screen of cDNA libraries	65
2.9.2. Direct two-hybrid analysis.....	66
2.10. Immunoprecipitation	66
2.11. Detection of snRNAs	67
2.12. Purification of proteins.....	68
2.12.1. Purification of bulk SR proteins from HeLa cells.....	68
2.12.2. Purification of recombinant USSRp58.....	69
2.13. <i>In vivo</i> splicing assays.....	70
2.14. <i>In vitro</i> splicing assays.....	71
2.14.1. Preparation of transcripts.....	71
2.14.2. Immunodepletion of HeLa nuclear extracts.....	71
2.14.3. <i>In vitro</i> splicing.....	72
2.14.3. S100 complementation assays	73
Results.....	74
1.1.Characterisation of a novel RS domain-containing protein	74
1.1.1. Identification and cloning of USSRp58	74
3.1.2. USSRp58 orthologs and paralogs in other species	77
3.1.4. Subcellular localisation of USSRp58.....	82
3.1.6. Analysis of the expression of USSRp58	87
3.1.7. USSRp58 interacts with splicing factors in two-hybrid analysis	91
3.1.10. USSRp58 does not complement splicing-deficient S100 HeLa extracts	100
3.1.12. USSRp58 can regulate alternative 5' site selection <i>in vivo</i> and <i>in vitro</i>	103
3.2. Identification of a nuclear retention signal in the RS domain of non-shuttling SR proteins.....	106
Discussion.....	110

4.1. Structural features of USSRp58	111
4.2. Expression and subcellular localisation of USSRp58.....	113
4.3. Interactions between USSRp58 and spliceosomal components.....	115
4.4. The role of USSRp58 in splicing.....	117
4.5. Nuclear retention signals in the RS domains of non-shuttling SR proteins.....	121
4.6. Future perspectives	122
References	125

Table of Figures

Figure 1: Two-step chemical mechanism for pre-mRNA splicing	14
Figure 2: Conserved sequences in nuclear pre-mRNA introns and self-splicing group II introns	15
Figure 3: The alternative splicing patterns that can be used by multiexon-containing pre-mRNAs	19
Figure 4: Spliceosome assembly pathway	22
Figure 5: Schematic diagram of human SR proteins	26
Figure 6: SR proteins are involved at numerous steps of pre-spliceosome assembly	37
Figure 7: Schematic diagram of SR-related proteins	42
Figure 8: Identification of a novel RS domain-containing protein	75
Figure 9: Sequence alignment of the human USSRp58 protein with its homologues in other species	78
Figure 10: Purification of recombinant USSRp58 protein	80
Figure 11: Characterisation of anti-USSRp58 antibodies	81
Figure 12: Subcellular localisation of USSRp58	83
Figure 13: Analysis of nucleocytoplasmic shuttling of USSRp58 by transient transfection in interspecies heterokaryons	86
Figure 14: Northern blot analysis of mouse USSRp58	88
Figure 15: Expression of USSRp58 in tissues and cultured cells	90
Figure 16: USSRp58 interacts with RS domain-containing proteins in a yeast two-hybrid system	92
Figure 17: USSRp58 interacts with splicing factors in a yeast two-hybrid system	95
Figure 18: USSRp58 interacts with splicing factors in cultured mammalian cells	97
Figure 19: Immunoprecipitation of snRNPs associated with USSRp58	99
Figure 20: USSRp58 does not complement splicing-deficient S100 extracts and can regulate 5' splice site selection <i>in vitro</i>	101
Figure 21: Effect of USSRp58 immunodepletion on pre-mRNA splicing <i>in vitro</i>	102
Figure 22: USSRp58 can regulate 5' splice site selection <i>in vivo</i>	104
Figure 23: C-terminal deletions in the RS domain of SC35 relieve nuclear retention	108
Figure 24: The NRS sequence identified in the RS domain of SC35 is transferable	109

Declaration

I declare that:

1. this thesis was composed by myself.
2. that the work presented is of my own unless otherwise stated.
3. that the results presented in section 3.2. and included in figures 23 and 24, have been published in Cazalla et al., (2002) Molecular and Cellular Biology 22(19): 6871-6882

Demian Cazalla

21/10/03

Date

Acknowledgements

I am grateful to Javier Cáceres, my supervisor, for inspiration, guidance, and support throughout my PhD. I would also like to thank him for comments and constructive criticism during the preparation of this thesis. I would also like to thank my second supervisor, Wendy Bickmore, for being helpful in the preparation of this thesis and in many other occasions.

I would like to thank to the other members of the Cáceres lab for help in everything. In particular, I would like to thank Jeremy Sanford. His comments and technical assistance were of invaluable help to complete this Ph.D. I would like to mention my gratitude to Janet Patridge and Graham Dellaire for help in my early days in the lab. I would also like to say thanks to the other post-docs in Bickmore's lab, specially Heidi Sutherland and Nick Gilbert, for constant technical support and advice. People from Mary O'Connell's lab have assisted me in countless occasions with reagents and technical help. Thanks Gillian for checking my English throughout the thesis!

I would like to mention Sandy Bruce and Douglas Stuart, from the Photography Department, for help in the preparation of the figures. Phillipe Gautier and Colin Semple assisted me with the bioinformatic analysis presented in this thesis

Abstract

Using a modified gene trap strategy, a novel factor involved in mRNA processing, USSRp58, was identified. Analysis of the USSRp58 sequence revealed that it contains an RS domain, common to factors involved in pre-mRNA splicing, and more generally in RNA processing. The subcellular localisation of the endogenous USSRp58 protein, as well of transiently expressed epitope-tagged USSRp58 protein, revealed that this protein localises to nuclear speckles, which suggests a potential role in pre-mRNA splicing. Two-hybrid analysis and immunoprecipitation/Western blot assays have been used to demonstrate that USSRp58 interacts with different proteins that have either potential or clearly defined functions in splicing. Immunoprecipitation experiments revealed that USSRp58 also interacts with spliceosomal snRNPs, indicating that it is physically associated with the splicing apparatus. Using *in vivo* and *in vitro* splicing assays, USSRp58 was shown to be functionally involved in splicing. Transient overexpression of USSRp58 in cultured cells and analysis of the alternative splicing of an adenovirus E1A reporter indicated that the protein can regulate 5' splice site selection *in vivo*. Furthermore, nuclear extracts immunodepleted of USSRp58 are unable to perform the second step of splicing *in vitro*, suggesting that this protein is necessary for this step.

A final set of results describing a nuclear retention signal (NRS) in the RS domain of SC35 was also included in this thesis. Chimeric proteins in which the RS domain of a non-shuttling SR protein, SC35, was fused to either hnRNP A1 or SF2/ASF (both shuttling proteins), were used to demonstrate that an NRS was present in the RS domain of SC35. Analysis of deletion mutants of SC35 indicated that the NRS is comprised in the last 30 amino acids of the protein. Finally, fusion of the NRS found in SC35 to SF2/ASF showed that this NRS is both portable and dominant.

Abbreviations used

A	adenine
Ac	acetate
ActD	actinomycin D
Ade	adenine
bp	base pairs
BSA	bovine serum albumine
cDNA	complementary deoxyribonucleic acid
C	cytosine
CHX	cycloheximide
Ci	Curies
CMV	cyto megalovirus
C-terminal	carboxy-terminal
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanine triphosphate
dNTP	deoxynucleotide triphosphate
ds	double strand
dTTP	deoxythymidine triphosphate
dH ₂ O	distilled water
Da	Daltons
DAPI	4,6-diamino-2-phenylindole
DMEM	Dulbecco's modified eagle's medium
DNA	deoxyribonucleic acid
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetra-acetic acid
ES	embryonic stem
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
EtOH	ethanol
FISH	fluorescence <i>in situ</i> hybridisation
FITC	fluorescein isothiocyanate
g	grams
G	guanine
GFP	green fluorescent protein
His	histidine
hnRNA	heterogeneous ribonucleic acid
hnRNP	heterogeneous ribonucleoprotein particle
HRP	horse radish peroxidase

hrs	hours
IgG	immunoglobulin G
IP	immunoprecipitate
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
k	kilo
kb	kilobases
KCl	potassium chloride
kDa	kilodaltons
KLH	keyhole limpet hemocyanin
l	litre
Leu	leucine
m	milli
μ	micro
M	molar
mAb	monoclonal antibody
Mb	megabases
min	minutes
MgCl ₂	magnesium chloride
mRNA	messenger RNA
mRNP	messenger ribonucleoprotein particle
n	nano
NaCl	sodium chloride
NaOH	sodium hydroxide
NLS	nuclear localisation signal
NRS	nuclear retention signal
N-terminal	amino-terminal
NTP	nucleotide triphosphate
OD	optical density
oligo	oligonucleotide
o/n	overnight
p	pico
P	phosphorus
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
pCp	Cytidine 3',5'-bis(phosphate)
PCR	polymerase chain reaction
PEG	polyethylene glycole
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute

rRNA	ribosomal ribonucleic acid
RT	room temperature
RT-PCR	reverse-transcriptase polymerase chain reaction
SD	synthetic dropout medium
SDS	sodium dodecyl sulphate
sec	seconds
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein particle
ss	single strand
TE	tris-EDTA
tRNA	transfer ribonucleic acid
Trp	tryptophan
TxRD	Texas Red
UTR	untranslated region
UV	ultraviolet
vol	volume
X- α -gal	5-Bromo-4-Chloro-3-indolyl- α D-galactopyranoside
YPD	yeast complete medium

Introduction

1.1. Gene expression and RNA processing

In order to make proteins, living organisms must copy the genetic information from a DNA template into an RNA transcript. In eukaryotes, these RNA transcripts must undergo three major processing events in the nucleus before they can be transported to the cytoplasm, where they will be recognized by the translation machinery as fully translatable mRNAs.

In eukaryotes, gene expression begins with transcription of protein-encoding genes by RNA polymerase II (RNA pol II) to make mRNA precursors (pre-mRNAs). The nascent pre-mRNA acquires a cap structure at the 5' terminus by the capping machinery, non-coding intervening sequences called introns must be removed through a process called splicing, and the 3' end of the transcript must be cleaved and polyadenylated. The mature mRNA can be then released from the site of transcription and exported to the cytoplasm for translation. Distinct machines are responsible for each of these processes, and this sequence of processing events, also includes an RNA surveillance system that ensures that aberrantly processed pre-mRNAs are eliminated.

It was originally considered that each of these processes occurred post-transcriptionally and were independent of each other. However, over the past few years, reports from several laboratories have suggested that these reactions can influence each other and that they are tightly linked to transcription. It is now believed that most RNA

processing events occur co-transcriptionally (for recent reviews see Bentley, 2002; Maniatis and Reed, 2002; Proudfoot et al., 2002).

This thesis will focus on the functional characterisation of trans-acting factors involved in pre-mRNA splicing.

1.2. Splicing

The discovery that eukaryotic genes are interrupted by blocks of sequence that are not used for producing proteins occurred more than 25 years ago (Berget et al., 1977; Chow et al., 1977). When analysed on the electron microscope, bulges were observed on RNA/DNA heteroduplexes prepared by hybridisation of adenovirus genomic DNA and cellular RNAs prepared from cells infected with the virus. Segments of sequence that were found together at the mRNA level, were shown to be interrupted at the DNA level. This result implied the existence of a process that should be in charge of removing the intervening sequences that were not found in the cytoplasmic mRNA. This process was named splicing.

The genes of all eukaryotic organisms contain these intervening non-coding sequences, called introns, that need to be removed in order to join together the blocks of coding sequence, called exons, to generate a functional mRNA. The complexity of the process and number of spliced genes increases with the complexity of the organisms. In the yeast *Saccharomyces cerevisiae* only about 250 genes contain introns, and most of these intron-containing genes have only one small intron. In humans, more than 99% of the genes are interrupted by introns, and the number and size of these vary enormously.

The splicing reaction is carried out by a macromolecular complex, the spliceosome. This megacomplex consist of five uridine-rich small nuclear ribonucleoprotein complexes

U1, U2, U4, U5 and U6 snRNPs, and a large number of non-snRNP associated proteins. The spliceosome is a highly dynamic complex and acts through a multitude of RNA-RNA, RNA-protein and protein-protein interactions. The RNA components of the snRNPs align the pre-mRNA splice sites and probably mediate splicing catalysis (Valadkhan and Manley, 2001). The associated proteins required for splicing mediate the structural rearrangements during spliceosome assembly and catalysis of the reaction. They also mediate the recognition of the signals present in the pre-mRNA molecule that are necessary for splicing (Hastings and Krainer, 2001; Staley and Guthrie, 1998).

The chemical reaction catalysed by the spliceosome consist of two phosphoryl transfer reactions (Fig. 1). Even though the spliceosome is an energy-consuming machine, no NTP hydrolysis is required for the splicing reaction . In the first step, the 2' hydroxyl of a conserved intronic adenosine attacks the phosphate at the 5' splice site, producing a free 5' exon and a branched molecule, termed the lariat intermediate. In the second step, the 3' hydroxyl group of the 5' exon attacks the phosphate at the 3' splice site, generating the ligated mRNA and a lariat intron, which is rapidly degraded (reviewed in Hastings and Krainer, 2001).

In order to perform the reactions described above, the spliceosome has to recognise the exons and introns in the pre-mRNA molecule. Signals in the pre-mRNA help to direct the assembly of the spliceosome for each splicing event. The borders of an intron are defined by 5' and 3' splice sites. In mammals, the 5' splice site consensus is AG/GURAGU (the / denotes the exon/intron boundary). The 3' splice site is defined by three different sequence elements: the branchpoint sequence, YNYURAC, located 18-40 nucleotides upstream the 3' splice site; the polypyrimidine tract; and finally the actual 3' splice site consensus YAG/N (Burge et al., 1999) (Fig. 2). A splicing error that adds or removes one nucleotide will disrupt the open reading frame of an mRNA. Since the sequences that

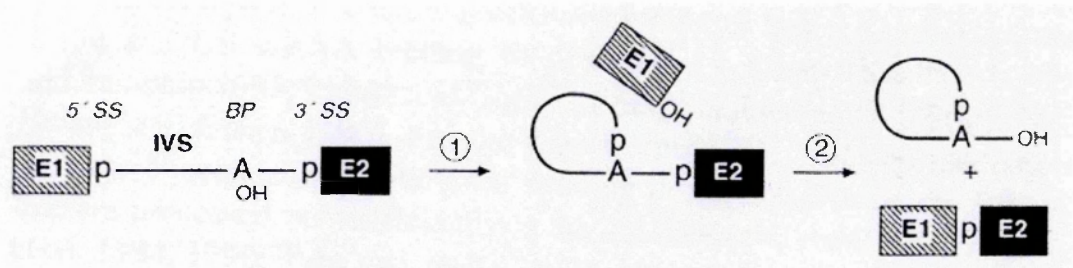
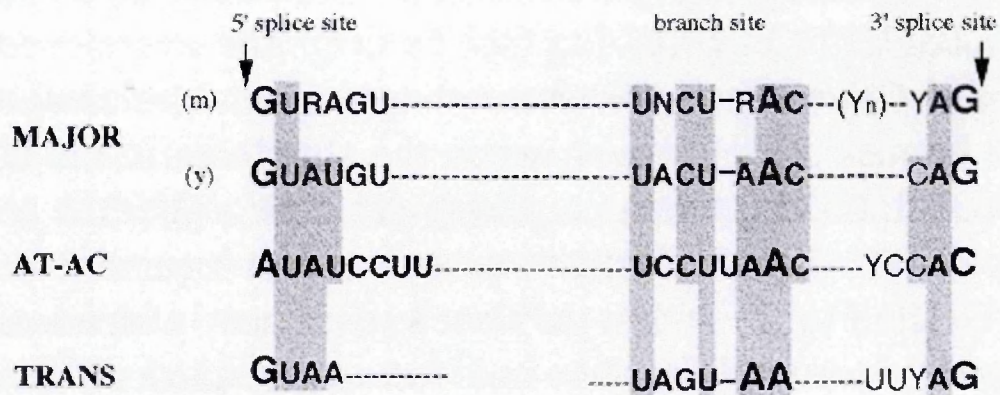


Figure 1: Two-step chemical mechanism for pre-mRNA splicing. A pre-mRNA with a single intron is shown at left, with exons (E1, E2) shown as boxes and the intron (IVS) shown as a line. The phosphodiester linkages that are broken or formed during the reaction are represented by the letter *p*. The branch adenosine (A) and 2' and 3' hydroxyl groups (OH) are also indicated. The ligated exon product and released lariat are shown at right. (Taken from Burge et al., 1999)

Pre-mRNA introns



Group II introns

GUGYG-----A-----AY

Figure 2: Conserved sequences in nuclear pre-mRNA introns and self-splicing group II introns. The 5' and 3' splice-site and branch-site consensus sequences are shown (R: purines; Y: pyrimidines). The *trans* branch-site sequence is from a nematode *trans*-splicing substrate. Enlarged letters in each intron indicate highly conserved nucleotides. Y_n represents a long stretch of pyrimidines that is present in mammalian (m) introns; group II and most yeast (y) introns do not contain such a sequence; nematode *trans*-splicing substrates usually have a short stretch of pyrimidine residues. Identical nucleotides in major-class introns, AT-AC introns, and *trans*-splicing introns are shaded. (Taken from Yu et al., 1999)

define introns and exons and short and degenerate, the spliceosome faces a major problem when it has to find the correct splice sites and splice exons correctly from within tens of thousands of intronic nucleotides. Auxiliary *cis*-elements, known as splicing enhancers that can be either exonic (ESEs) or intronic (ISEs), facilitate the recognition of exons. Exon recognition is then accomplished by the accumulated recognition of weak signals. This results in a network of interactions across exons as well as across introns (Berget, 1995; Reed, 1996), between splicing factors that bind to these weak signals. It is also clear that different exons are recognised by different mechanisms and require different sets of auxiliary elements in addition to the common splice-site sequences.

The presence of the cap binding complex (CBC) has been shown to increase the efficiency with which U1 snRNP binds to the cap-proximal 5' splice site and thus increase the rate of recognition and splicing of the cap-proximal intron (Colot et al., 1996; Lewis et al., 1996). In terminal exons, other elements can help the spliceosome to define the boundaries. The polyadenylation signal has been shown to promote the use of the last 3' splice site (Lewis et al., 1995).

It has been demonstrated that the elongation rate at which RNA pol II transcribes a DNA template can influence the efficiency of recognition of a weak splice site by the spliceosome (Kadener et al., 2001; Kadener et al., 2002; Nogues et al., 2002). Therefore, the kinetics of transcription can also help to define which splice sites have to be used by the spliceosome, and this is most relevant in the context of alternative splicing (see section 1.3).

In lower organisms, including trypanosomatid protozoans, nematodes, trematodes, and euglenoids, exons from separately transcribed pre-mRNAs can be spliced by a process called *trans*-splicing (reviewed in Nilsen, 1995). In this process, exons are spliced intermolecularly, rather than from within a single precursor RNA as in conventional *cis*-

splicing. In this reaction, which is catalysed by another type of spliceosome, the 5' exon is supplied by a small RNA termed the splice leader RNA (SL RNA), which has similarities to the snRNAs that participate in *cis*-splicing. In nematodes, the SL RNA interacts with specific proteins to form the spliced leader ribonucleoprotein (SL RNP), which is then competent to work in *trans*-splicing (Maroney et al., 1990). In addition to the SL RNP, other factors including U2, U4/U6, and U5 are also required (Hannon et al., 1991; Maroney et al., 1996). Splicing is not only necessary to accurately generate the message that will be translated into a protein, but also influences subsequent steps of RNA processing, such as mRNA export. A multi-protein complex is deposited on the mRNA in a sequence-independent, position-dependent manner, 20-24 nucleotides upstream of mRNA exon-exon junctions (EJC, for exon junction complex). The EJC is necessary for recognition of the mRNA by the mRNA transport machinery (Le Hir et al., 2000; Luo and Reed, 1999), and also plays a role in the fate of the mRNA in the cytoplasm (reviewed by Dreyfuss et al., 2002).

1.3. Alternative splicing

Alternative splicing is the joining of different 5' and 3' splice sites, allowing the cell to increase protein diversity by the generation of multiple, sometimes functionally distinct, protein isoforms to be encoded by a single gene. Up to 59% of human genes generate multiple RNAs by alternative splicing (Lander et al., 2001), and approximately 80% of alternative splicing results in changes in the encoded protein (Modrek and Lee, 2002). This is supposed to be a low estimate, since ESTs derive from a limited number of tissues or developmental states, and cover only a limited portion of each mRNA. These facts

postulate alternative splicing as the primary source of human proteomic diversity. There are remarkable examples of hundreds or even thousands of functionally different mRNAs or proteins produced from a single gene. One of such examples is the *Drosophila DSCAM* gene, which codes for a protein that works as an axon guidance receptor responsible for directing growth cones to their proper target in Bolwig's nerve of the fly (Schmucker et al., 2000). Its pre-mRNA is alternatively spliced and can potentially generate 38016 different protein isoforms (reviewed by Graveley, 2001). The *Drosophila* genome contains approximately 13600 identified genes (Adams et al., 2000), whereas a single gene can produce nearly as much as three times that number of proteins. These facts postulate alternative splicing as the primary source of human proteomic diversity

Alternative splicing occurs when the splicing machinery has the possibility to choose from more than one set of splice sites, involving competition among potential splice sites. Splice sites that are usually recognised by the spliceosome as weak signals, can be positively regulated by *trans*-acting factors that bind to splicing enhancers. On the other hand, strong sites can be negatively regulated by *trans*-acting factors, making weak signals a better option for the spliceosome (Caceres and Kornblihtt, 2002). The spliceosome can face situations in which it can use alternative 5' splice sites, alternative 3' splice sites, optional exons, mutually exclusive exons, or it can let an intron be retained in the final transcript (Fig. 3). Some alternative splicing events seem to be constitutive, with different mRNA isoforms coexisting with a constant relative abundance in the cell, where other events are regulated, producing different isoforms in different cell types or at different developmental stages.

The specificity seems to be achieved by the relative concentration or activity of competing and cooperating factors together with the strength of the splice sites and the strength of the binding sites for the *trans*-acting factors. Regulation is then usually

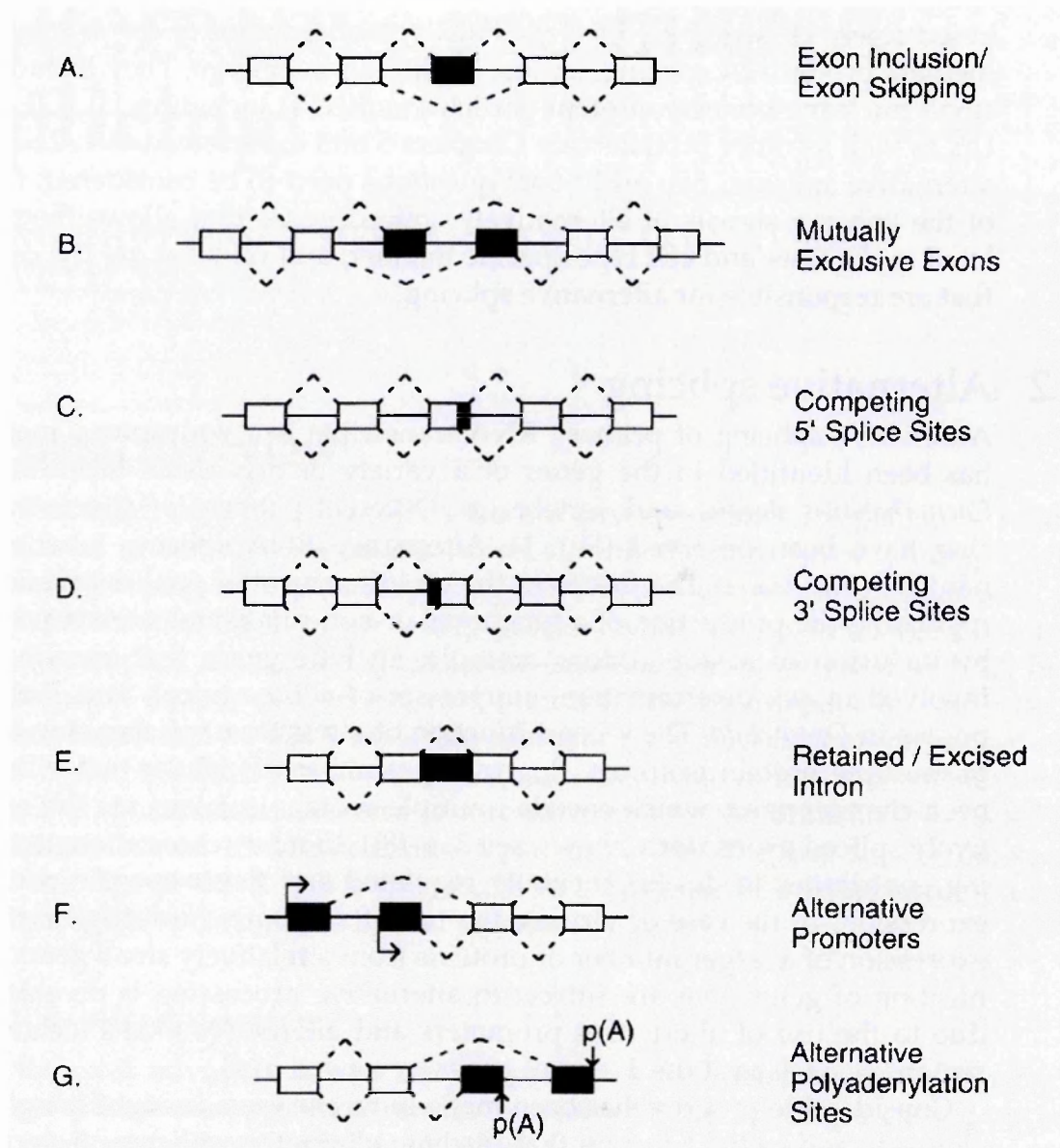


Figure 3: The alternative splicing patterns that can be used by multiexon-containing pre-mRNAs. Boxes represent exons and lines represent introns. Open boxes represent the constitutively expressed exons, whereas solid boxes represent the alternatively spliced exons. Dash lines represent splicing pathways. In f, the arrows indicate the alternative transcription initiation sites, which define the 5' boundaries of the alternative first exons. In g, the arrows indicate the alternative sites of 3' cleavage and polyadenylation, which define the 3' boundaries of the alternative last exons. More complex combinations of patterns a-g are found in complex transcription units. (Taken from Wang et al., 1997)

accomplished in the early events in the assembly of the spliceosome on the splice sites (Smith and Valcarcel, 2000). Nevertheless, there are examples of spliceosomal proteins, like SPF45, that interact with the AG at the 3' splice site and can activate alternative 3' splice sites during the second step of splicing (Lallena et al., 2002).

1.4. The spliceosome

The spliceosome consists of the five small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4, U5 and U6, and 50-100 non-snRNP splicing factors (Kramer, 1996). Each snRNP particle consists of a multi-subunit complex between an uracyl-rich small nuclear RNA (U snRNA) and several proteins, some of which are unique to the individual snRNP and others that are common to many spliceosomal snRNPs (Sm or Sm-like proteins) (Will and Luhrmann 2001).

The spliceosomal snRNPs are very abundant, with over a million copies per cell. These snRNPs are part of the major spliceosome, that is responsible for splicing the vast majority of pre-mRNA introns (so-called U2-type introns). There is also a set of less abundant snRNPs, U11, U12, U4atac and U6atac, that together with U5, are subunits of the so-called minor spliceosome that splices a rare class of pre-mRNA introns, denoted U12-type (reviewed by Burge et al., 1999). The consensus sequences recognised by the minor spliceosome are different from those recognised by the major spliceosome, and are more strongly conserved than those of U2-type introns. The minor spliceosome splices introns containing either /ATATCCTTT and YAC/ or /GTATCCTTT and YAG/, respectively (where / represents the splice junction). For both the AT-AC and GT-AG subtypes of this rare intro class, the conserved branch site signal follows the consensus TCCTTAAC

(branch site underlined). The proportion of U12-type introns is very small, about one in a thousand.

U snRNP biogenesis is a complex process (for a review see Will and Luhrmann, 1997). With the exception of U6, all the spliceosomal snRNAs are transcribed by RNA pol II as snRNA precursors that contain additional 3' nucleotides and acquire a monomethylated, m7GpppG (m7G) cap structure. Then, these pre-U snRNAs are exported to the cytoplasm, where they interact in an ordered, stepwise manner with seven Sm proteins, B/B', D3, D2, D1, E, F, and G, to form the snRNP Sm core structure. Subsequently, the m7G cap is converted to the 2,2,7-tri-methylated guanosine (m3G), and U snRNAs undergo 3' end maturation. After processing in the cytoplasm, U snRNP are re-imported to the nucleus, where they assemble with their individually specific snRNP proteins. Although less is known about the maturation process of the minor U11, U12 and U4 atac snRNPs, it is assumed that they follow a pathway similar to that described for the major spliceosome snRNPs. Biogenesis of U6 snRNP, and presumably U6atac snRNP, differs in many ways from the one described for the other spliceosomal snRNPs. The U6 snRNA is transcribed by RNA pol III and the processing occurs entirely in the nucleus. The U6 snRNA acquires a γ -monomethyl cap structure and is also internally modified primarily by pseudouridylation and 2'-O-methylation. The U6 snRNA assembles with Sm-like proteins. Finally, the mature U6 snRNP particle pairs to U4 snRNP to form the U4/U6 snRNP.

The spliceosome assembles in a stepwise fashion (Fig. 4) (for a review see Staley and Guthrie, 1998). The commitment of pre-mRNA to the splicing pathway occurs upon ATP-independent formation of the E (early) complex. Assembly of the E complex involves recognition of the 5' splice site by U1 snRNP base pairing, binding of SF1/mBBP (splicing factor 1/ mammalian branch point binding protein) to the branchpoint sequence and

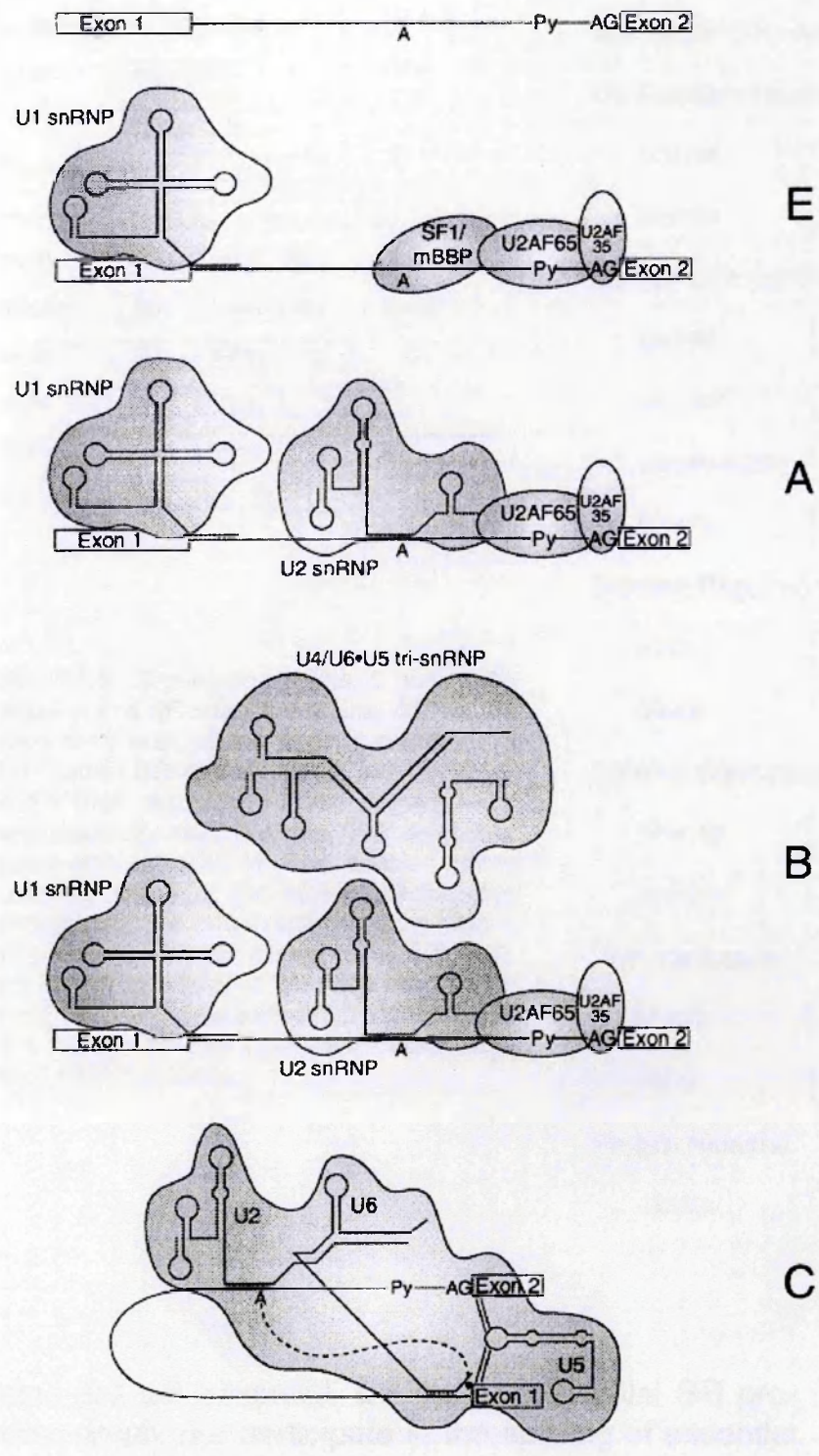


Figure 4: Spliceosome assembly pathway. See text for description.
(Taken from Graveley, 2000)

association of the U2 auxiliary factor (U2AF), which binds to the polypyrimidine tract and the AG at the 3' splice site. Next, U2 snRNP base pairs with the branchpoint sequence during the ATP-dependent formation of the A-complex. The subsequent association of the U4/U6-U5 tri-snRNP with the pre-mRNA results in the formation of the B complex, and finally, the C complex is formed by rearrangements of RNA-RNA, RNA-protein and protein-protein interactions to create the catalytically competent spliceosome. Members of the serine and arginine-rich proteins (SR proteins) function as molecular adapters, mediating interactions between the pre-mRNA and the assembling spliceosome throughout the whole pathway. Their function and characteristics will be discussed later in this chapter.

The recognition of the 5' splice site by U1 snRNP involves base pairing between the 5' splice site and 10 strongly conserved nucleotides at the 5' end of U1 snRNA; these are highly complementary sequences (Seraphin et al., 1988; Siliciano and Guthrie, 1988; Zhuang and Weiner, 1986). In the absence of base pairing, the U1 snRNP U1C protein recognises the 5' splice site (Du and Rosbash, 2002). Although the presence of U1 snRNP is necessary for splicing *in vitro*, increasing amounts of SR proteins can restore splicing in nuclear extracts depleted of U1 snRNP by antisense affinity depletion (Crispino et al., 1994; Tarn and Steitz, 1994). In this case, the 5' splice site is recognised by U6 snRNP (Crispino and Sharp, 1995). The binding of U1 snRNP is destabilised when the U4/U6-U5 tri-snRNP particle is added to the spliceosome, followed by several rearrangements in which U1 is replaced by U5 and U6 at the 5' splice site (Kandels-Lewis and Seraphin, 1993).

U2 snRNP associates with the E complex in an ATP-independent manner that does not require the branch point sequence, but its stable binding to the branch point sequence requires ATP and several proteins (Das et al., 2000). U2AF, an essential splicing factor purified as a heterodimer composed of 65 kDa (U2AF⁶⁵) and 35 kDa (U2AF³⁵) subunits,

facilitates U2 snRNP binding to the pre-mRNA (Zamore and Green, 1989). U2AF⁶⁵ binds directly to the polypyrimidine tract, whereas U2AF³⁵ binds to the AG at the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). Several studies support a mechanism whereby SR proteins bound to splicing enhancers interact with U2AF³⁵, thereby recruiting U2AF⁶⁵ to a weak 3' splice site (reviewed in Graveley, 2000). SF1/mBBP cooperatively binds to U2AF⁶⁵ to promote the binding of U2 snRNP to the branchpoint sequence (Berglund et al., 1998), a process that involves other proteins as UAP56, p54 and SAP155 (Fleckner et al., 1997; Gozani et al., 1998; Zhang and Wu, 1996). Other proteins that are also required for stable binding of U2 snRNP to the branch point sequence are SF3a and SF3b, which comprise multiple subunits (Brosi et al., 1993a; Brosi et al., 1993b). The SF3b protein subunit p14, cross-links to the branch point adenosine after the incorporation of the U2 snRNP into the spliceosome, and it can be cross-linked in subsequently formed complexes, including the catalytically active spliceosomal C complex (MacMillan et al., 1994; Query et al., 1996).

Upon the joining of the U4/U6-U5 tri-snRNP, several rearrangements of RNA-RNA interactions take place to form the active spliceosome. U5 snRNP makes contact with the 5' and 3' ends of the intron. This interaction brings the two splice sites together. A U5 snRNP-specific protein, Prp8 (U5-220 kDa, p220 or hPrp8) is closely associated with the sequences at the 5' splice site, branch site, and 3' splice site, implying a function as a major scaffolding protein for aligning the splice sites. U6 snRNP dissociates from U4 snRNP and interacts with U2 snRNP. Upon interaction with U2 snRNP, U6 snRNP displaces U1 snRNP and remains bound to the 5' splice site of the intron (Murray and Jarrell, 1999), and this change positions the branch site adenosine for attack at the 5' splice site (Madhani and Guthrie, 1994). The lariat-exon intermediate and the free 5' exon are formed during the first catalytic step of splicing. Before the second step of splicing, additional rearrangements

occur, creating new RNA-RNA interactions. The U2-U6 snRNP interaction changes and contacts the 5' splice site in the lariat intermediate and the free 5' exon is joined together with the 3' exon (Burge et al., 1999). Several DExD/H box ATPases are believed to function as motors that drive the RNA rearrangements during splicing (Staley and Guthrie, 1998).

This idea of an order of addition of discrete U1, U2, and U4/U6-U5 snRNPs to the pre-mRNA substrate during spliceosome assembly has been challenged by the recent discovery in yeast extracts of a very large particle containing all five of the U snRNPs. The existence of this penta-snRNP particle suggests that the spliceosome may actually exist as a pre-assembled multi-snRNP particle, independently of a pre-mRNA substrate (reviewed in Nilsen, 2002; Stevens et al., 2002).

1.5. SR proteins

The SR proteins are a highly conserved family of essential pre-mRNA splicing factors (for reviews see Fu, 1995; Graveley, 2000; Manley and Tacke, 1996). They are required for constitutive splicing as well as for the regulation of alternative splicing. SR proteins have a modular domain structure consisting of one or two copies of an N-terminal RNA recognition motif (RRM, also known as RBD, for RNA-binding domain) and a C-terminal domain rich in alternating serine and arginine residues, known as RS domain (Fig. 5). The RRM determines the specificity in binding to RNA, while the RS domain participates in protein-protein interactions.

SR proteins were independently discovered by a number of groups taking very different approaches. The prototypical SR protein, SF2/ASF (splicing factor 2/alternative

A. Human SR Proteins

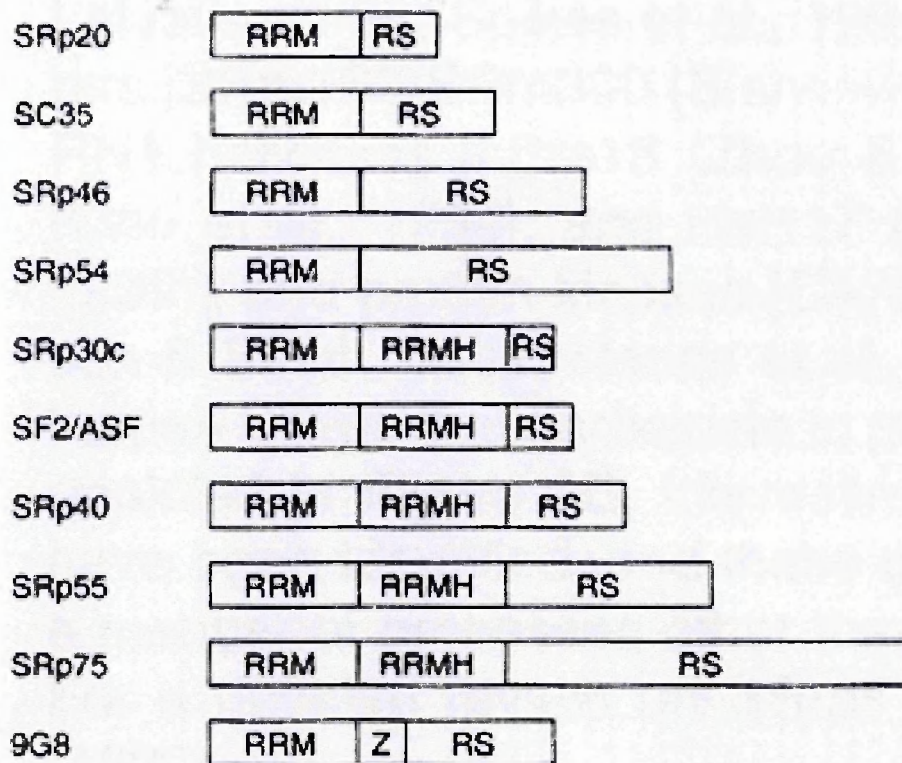


Figure 5: Schematic diagram of human SR proteins. The domain structures of the known members of the human SR protein family are depicted. RRM: RNA recognition motif; RRMH: RRM homology; Z: zinc knuckle; RS: arginine/serine-rich domain. (Taken from Graveley, 2000)

splicing factor, also referred to as ASF/SF2), was discovered and characterised independently by two different groups employing two different assays. ASF was purified as an activity that could regulate the selection of alternative 5' splice sites (Ge and Manley, 1990), suggesting a role for ASF in alternative splicing. On the other hand, SF2 was identified initially as an activity required for splicing of a β -globin pre-mRNA *in vitro* (Krainer and Maniatis, 1985) and was subsequently purified as a factor essential for splicing (Krainer et al., 1990b), but with the ability to modulate alternative splicing *in vitro*, in a concentration-dependent manner (Krainer et al., 1990a). Posterior isolation of cDNAs encoding the proteins confirmed that they were identical (Ge et al., 1991; Krainer et al., 1991). SC35 was identified with a monoclonal antibody raised against purified spliceosomes and was shown to be necessary for splicing and spliceosome assembly *in vitro* (Fu and Maniatis, 1990; Fu and Maniatis, 1992). Most other SR proteins were identified based on their reactivity with a monoclonal antibody, mAb104, that recognises a phosphorylated epitope and reacts against active sites of RNA pol II transcription on lampbrush amphibian chromosomes (Roth et al., 1991). All the proteins recognised by mAb104 could be easily purified by a simple two-step purification involving selective precipitation in the presence of millimolar concentrations of $MgCl_2$ (Zahler et al., 1992). By following this two-step protocol it is possible to obtain homogeneous preparations consisting of six major proteins ranging in apparent size from 20 to 75 kDa. Proteins of similar sizes could be purified from different species, suggesting that the protein family is evolutionary well conserved from *Drosophila* to humans. Partial amino acid sequence revealed that all the proteins shared sequence similarity, and those of approximately 30 kDa were found to be identical to SF2/ASF and SC35. The purified proteins also all have the ability to function as essential splicing factors. Roth and colleagues (Zahler et al., 1992) proposed five criteria that define the family of SR proteins: (1) They contain a shared

phosphoepitope recognised by the monoclonal antibody mAb104. (2) They copurify in a two-step salt precipitation procedure (soluble in 65% ammonium sulfate and precipitated in 20 mM MgCl₂). (3) They share a similar protein structure containing at least one RRM at the N-terminus and an RS domain at the C-terminus. (4) Their sizes on SDS-PAGE are conserved from *Drosophila* to man. (5) They can complement splicing deficient S100 extracts. Degenerate primers matching conserved features in the RNP-2 and RNP-1 submotifs of RRM1, or the conserved heptapeptide present in RRM2 allowed the identification and cloning of more members of the family (Kim and Baker, 1993; Kim et al., 1992; Screaton et al., 1995). One of the proteins identified by this method, SRp30c, does not fully comply with the requirements proposed by Roth and colleagues to be assigned to the family, since it was not detected after purification with the two-step protocol, probably due to low abundance. Another member of the family, p54, was isolated by expression cloning using an anti-lamin B autoantiserum and is similar in size and structure to SRp75 (Chaudhary et al., 1991). Although the protein is less homologous to the canonical SR proteins, it has been shown to be reactive to mAb104 and to be capable of complementing S100 extracts (Zhang and Wu, 1996). Finally, the nuclear antigen 9G8 was identified by a monoclonal antibody raised against heterogeneous nuclear ribonucleoproteins (hnRNPs) (Cavaloc et al., 1994). Again, this protein does not fully comply with the definition of SR proteins, since it is recognised weakly by mAb104 and besides the N-terminal RRM and the C-terminal RS domain, it also contains a zinc finger motif. However, 9G8 is commonly considered as a member of the SR family. Although SR proteins have been identified in all metazoans examined and in plants, they are not present in all eukaryotes. Whereas two SR proteins have been identified in the fission yeast *Schizosaccharomyces pombe* (Gross et al., 1998; Lutzelberger et al., 1999), none has been found in *Saccharomyces cerevisiae*.

Additionally, there is a growing number of RS-containing proteins that are structurally very similar to bona fide SR proteins although they do not share all common functions. For example, SRrp86, a protein that contains a single RRM in its amino terminus and two carboxy-terminal RS domains separated by an unusual glutamic acid-lysine (EK)-rich region, is unable to complement splicing-deficient S100 extracts. Actually, SRrp86 inhibits the activity of bona fide SR proteins, such as SC35 or SF2/ASF, in an S100 complementation assay (Barnard and Patton, 2000). Likewise, SRrp30 and SRrp40 contain a single RRM and an RS domain, but are unable to complement S100 extracts and antagonise SR proteins in *in vitro* splicing assays (Cowper et al., 2001). In another report (Shin and Manley, 2002) SRrp40 (also called SRp38) was shown to repress general splicing in mitotic extracts (Shin and Manley, 2002).

The proteins described above belong to a group of proteins that are collectively known as SR protein related polypeptides (SRrp) or SR-related proteins. The only common feature of members of this group is the presence of an RS domain. SR-related proteins have diverse functions in the processing of pre-mRNA. A detailed description of the members of this group will be given later in this chapter (see section 1.8).

The SR proteins seem to be functionally redundant in constitutive splicing as illustrated by the ability of any individual SR protein to complement an otherwise inactive cytosolic HeLa S100 extract. However, several differences in the ability of these proteins to regulate alternative splicing, as well as the ability of individual SR proteins to commit different pre-mRNAs to the splicing pathway suggested that individual SR proteins may have unique functions in splicing regulation (reviewed in Sanford et al., 2003). Genetic approaches used to address this question showed that deletion of the gene encoding the B52 protein (ortholog of human SRp55) in *Drosophila melanogaster* results in lethality (Peng and Mount, 1995; Ring and Lis, 1994). Likewise, conditional depletion of SF2/ASF in

chicken DT40 cell line results in cell death (Wang et al., 1996). Interestingly, knockout of the SRp20 gene in the mouse blocked development at a very early stage (Jumaa et al., 1999). Additionally, Cre-mediated conditional deletion of SC35 in the thymus causes a defect in T cell maturation (Wang et al., 2001). These observations gave rise to the general impression that SR proteins are required for cell viability. However, when the expression of the genes encoding different SR proteins were ablated individually by RNA interference (RNAi) in *Caenorhabditis elegans*, no phenotype was observed for all the cases, except for the SF2/ASF ortholog, *rsp-3* (Longman et al., 2000). But when multiple *C. elegans* SR protein genes were simultaneously targeted by this method, developmental defects or lethality were observed (Longman et al., 2000). The basis for these results is still unknown. Either the nonessential SR proteins do simply not participate in the splicing of essential genes, or the other SR proteins present in the cell can functionally substitute for the missing one. In the case of the B52 knock-out in *Drosophila*, no general splicing defects were found in the null background, either for constitutive or alternatively spliced genes (Ring and Lis, 1994); however recent observations suggest that the lethality of the B52 deletion strain is a consequence of splicing defects in tissues in which B52 is normally the major SR protein (Hoffman and Lis, 2000).

1.6. The RNA recognition motif

The RNA recognition motif (RRM) is a conserved, modular domain of approximately 80 amino acids that contains two small, highly conserved sequence elements, the RNP-1 octamer and the RNP-2 hexamer (Hoffman et al., 1991; Nagai et al., 1990). The RRM motif is present in a variable number of copies in several RNA-binding proteins involved in

pre-mRNA and pre-rRNA processing (Birney et al., 1993; Kenan et al., 1991). Substitution of conserved residues within the RNP-1 and RNP-2 submotifs affects binding of the U1A and U1-70K proteins to U1 snRNA (Scherly et al., 1989; Surowy et al., 1989). SR proteins contain either one or two RRM. In those containing two RRMs, the second RRM is atypical, not highly conserved, and always contains the heptapeptide, SWQDLKD (Birney et al., 1993). Mutations in the RNP motifs of SF2/ASF inhibits binding of the protein to RNA, resulting in a decreased activity of the protein in pre-mRNA splicing (Caceres and Krainer, 1993). The RRMs of SR proteins bind RNA in a sequence-specific manner, even in the absence of the RS domain (Manley and Tacke, 1996). SELEX protocols (selected evolution of ligands through exponential enrichment) have been employed to characterise the RNA binding specificity of individual SR proteins (Adams et al., 2000; Cavaloc et al., 1999; Liu et al., 1998; Schaal and Maniatis, 1999b; Tacke et al., 1997; Tacke and Manley, 1995). These studies led to the conclusion that SR proteins recognise a vast array of sequences (for a review, see Tacke and Manley, 1999). Although SR proteins do display distinct RNA binding specificities, the consensus sequences they recognise are rather degenerate.

The RRMs of SR proteins have also been suggested to mediate protein-protein interactions. Using chimeric proteins consisting of domain swaps between different SR proteins, it was shown that the specificity displayed by SR proteins in alternative splicing is determined by the RRMs (van Der Houven Van Oordt et al., 2000). In this study, the RRM2 of SF2/ASF was shown to have a dominant role in alternative splicing. One possible explanation for the dominant role of this domain in alternative splicing could lie in the establishment of new protein-protein interactions. In this scenario, the RRM2 of SF2/ASF could help to recruit additional splicing factors to the spliceosome and modulate splicing specificity. The RRM2 of SF2/ASF was also shown to be both necessary and sufficient to

repress splicing of the adenovirus IIIa pre-mRNA. (Dauksaite and Akusjarvi, 2002). The heptapeptide SWQDLKD, which is unlikely to be directly involved in RNA binding given its position in the predicted structure of RRM2, was essential for splicing repression. This observation suggested that the repressive activity shown by the RRM2 of SF2/ASF is achieved by directly interacting with either proteins that are required for early splice site recognition or that negatively affect splice site recognition.

1.7. The RS domain

The RS domain was originally identified in the *Drosophila* splicing regulators Tra, Tra2 and *suppressor-of-white-apricot* (SWAP). Subsequently, RS domains were also found in the U1 snRNP- specific, U1 70K protein, and in members of the SR protein family. The RS domain consists of simple arginine and serine repeats, that can be occasionally interrupted by other amino acids, and the length and sequence of RS domains are usually conserved among species.

The RS domain of SR proteins mediates protein-protein interactions with other RS domain containing factors such as U1-70K and U2AF³⁵ (Amrein et al., 1994; Kohtz et al., 1994; Wu and Maniatis, 1993). However, not all RS-domain containing proteins can interact with one another, suggesting that different protein-protein interactions may have different RS domain sequence requirements. For example, SF2/ASF and SC35 interact with U1-70K and U2AF³⁵ (Wu and Maniatis, 1993), but p54 is unable to do so (Zhang and Wu, 1996). Moreover, SF2/ASF is able to interact with U1-70K, but its RS domain alone is not sufficient for this interaction. In contrast, the RS domain of SF2/ASF on its own is able to interact with RSF1, a *Drosophila* splicing repressor (Labourier et al., 1999).

Specific roles for the RS domains of individual SR proteins were suggested by the high phylogenetic conservation of specific sequences within the RS domains (Birney et al., 1993). Chimeric proteins containing different RS domains fused to the RRM of the bacteriophage MS2 protein activated splicing of substrates containing a single MS2 binding site, and this activation correlated directly with the number of RS dipeptides present in the RS domain, suggesting that different RS domains have unique activities (Graveley and Maniatis, 1998). Moreover, individual RS domains have unique properties in directing subcellular localization and influencing the ability of SR proteins to shuttle from the nucleus to the cytoplasm (Sanford et al., 2003). However, in other experimental systems, the RS domain of SR proteins seem to be functionally redundant. For instance, the RS domains of SF2/ASF and SC35 are functionally interchangeable in S100 complementation assays (Mayeda et al., 1999b). Interestingly, the RS domains are also functionally interchangeable *in vivo*, as shown by the ability of chimeric proteins consisting of the RRMs of SF2/ASF fused to RS domains of different SR proteins to rescue cell viability in a chicken B cell line DT40 (Wang et al., 1998). In contrast, it was recently shown that RS domains are not all functionally equivalent *in vivo* in *Drosophila* as different RS domains varied considerably in their ability to restore Tra2 function (Dauwalder and Mattox, 1998).

RS-domain containing proteins are generally widely distributed throughout the nucleus, excluding nucleoli, and are highly concentrated in subnuclear structures referred to as nuclear speckles, which correspond to the interchromatin granule clusters and associated perichromatin fibrils observed by electron microscopy (reviewed in Spector et al., 1993). There are about 20-40 speckle structures per nucleus, and they occupy ~20% of total nuclear volume. It has been proposed that the function of these speckles is the storage and/or assembly of spliceosomal components (Misteli, 2000). In addition, to pre-mRNA splicing factors and snRNAs, speckles also contain transcription factors (Zeng et al., 1997),

3'-processing factors (Schul et al., 1998), and ribosomal proteins (Mintz et al., 1999). Biochemical purification of these nuclear structures indicate that they contain about 150 proteins, although it remains to be determined which ones are bona fide residents (Mintz et al., 1999).

The subcellular localisation of SR proteins is influenced by the RS domain, which mediates the interaction with the SR protein nuclear import receptor, transportin-SR (Kataoka et al., 1999; Lai et al., 2000; Lai et al., 2001; Yun et al., 2003). Although the RS domain is a nuclear localisation signal, subnuclear targeting to the speckles requires at least two of the three constituent domains of SF2/ASF. In contrast, in two SR proteins that have a single RRM (SC35 and SRp20), the RS domain is both necessary and sufficient as a targeting signal to the speckles (Caceres et al., 1997; Hedley et al., 1995; Li and Bingham, 1991). The RS domain is also an important determinant of nucleocytoplasmic shuttling; however, this domain is not sufficient to promote shuttling of an unrelated protein reporter, suggesting that additional signals are required (Caceres et al., 1998).

1.8. SR protein phosphorylation

Whereas phosphorylated SR proteins are required for the assembly of the spliceosome (Mermoud et al., 1994), dephosphorylation of SR proteins is required for splicing catalysis (Xiao and Manley, 1998). The SR proteins domains are extensively modified by phosphorylation, mainly at the serine residues in the RS domain (Colwill et al., 1996b). Several kinases have been shown to phosphorylate RS domains. These include SRPK1 (Gui et al., 1994) and SRPK2 (Wang et al., 1998), the Clk-Sty family (Colwill et al., 1996a), DNA topoisomerase I (Rossi et al., 1996) and cdc2 (Okamoto et al., 1998).

DOA, an homologue of Clk/Sty in *D. melanogaster*, was shown to phosphorylate RBP1, Tra and Tra2 *in vitro* and RBP1 *in vivo* (Du et al., 1998) and to regulate the splicing of *dsx* pre-mRNA (Du et al., 1998). The phosphorylation of the RS domain appears to influence the subcellular localisation of SR proteins. Addition of purified SRPK1 to permeabilised cells, or overexpression of SRPK1, 2 or cdc2-like kinase (Clk)1 in transfected cells result in apparent disassembly of the nuclear speckles (Colwill et al., 1996a; Duncan et al., 1998; Gui et al., 1994; Kuroyanagi et al., 1998; Wang et al., 1998). These results suggest that phosphorylation, or hyperphosphorylation, causes release of these factors from the speckles, or perhaps that the integrity of these structures is compromised. Phosphorylation also causes redistribution of SR proteins from speckles to sites of transcription (Misteli et al., 1998). The phosphorylation of RS domains modifies their ability to interact with each other and with U1-70K (Wang et al., 1998; Xiao and Manley, 1998), and influences their ability to bind RNA (Tacke et al., 1997; Xiao and Manley, 1997). Since the state of phosphorylation of the RS domain of SR proteins affect their ability to interact with both proteins and RNA, it is expected that the state of phosphorylation of their RS domains would affect their activities in splicing. In fact, both phosphorylation and dephosphorylation of SR protein are necessary for splicing *in vitro* (Cao et al., 1997; Xiao and Manley, 1997; Xiao and Manley, 1998). Different steps in splicing require different extent of phosphorylation of the RS domain.

The level of phosphorylation of SR proteins also affect their activity in alternative splicing. SR proteins are highly phosphorylated *in vivo*, a modification that is required for inhibition the splicing of adenovirus IIIa pre-mRNA. It was shown that SR proteins purified from late adenovirus-infected cells are inactivated as splicing repressor proteins by virus-induced dephosphorylation (Kanopka et al., 1998). SR proteins are dephosphorylated by the endogenous protein phosphatase 2A, which is activated by the viral protein E4-

ORF4. E4-ORF4 binds directly to SF2/ASF, and this interaction seems to be necessary for relieving the repressive effect of SR proteins on IIIa pre-mRNA splicing in transient transfection experiments (Estmer Nilsson et al., 2001).

1.9. Functions of SR proteins

Roles in splicing

SR proteins show a great flexibility that allows them to function at numerous steps in spliceosome assembly and to regulate splice site selection (Fu, 1995; Graveley, 2000). The roles of SR proteins in splicing can be divided into two big categories: exon-dependent and exon-independent functions (Fig. 6) (Graveley, 2000).

Exon-dependent functions

In order to perform the exon dependent functions, SR proteins must bind to ESEs. Several papers showed that SR proteins could bind to ESEs to enhance the splicing of adjacent introns. For example, SR family proteins bound to ESEs can promote U2AF recruitment to the polypyrimidine tract and activate an adjacent 3' splice site (Graveley et al., 2001; Romfo et al., 2001). This activity involves the interaction between the RS domain of the ESE bound SR protein and the RS domain of U2AF³⁵ (recruitment model) (Zuo and Maniatis, 1996). But not all the cases of enhancer-dependent splice site selection can be explained by the recruitment model. In certain pre-mRNA substrates, the presence of ESE-bound SR proteins do not enhance the binding of U2AF to the polypyrimidine tract but rather helps to antagonise the activity of proteins bound to ESSs (Hastings and Krainer,

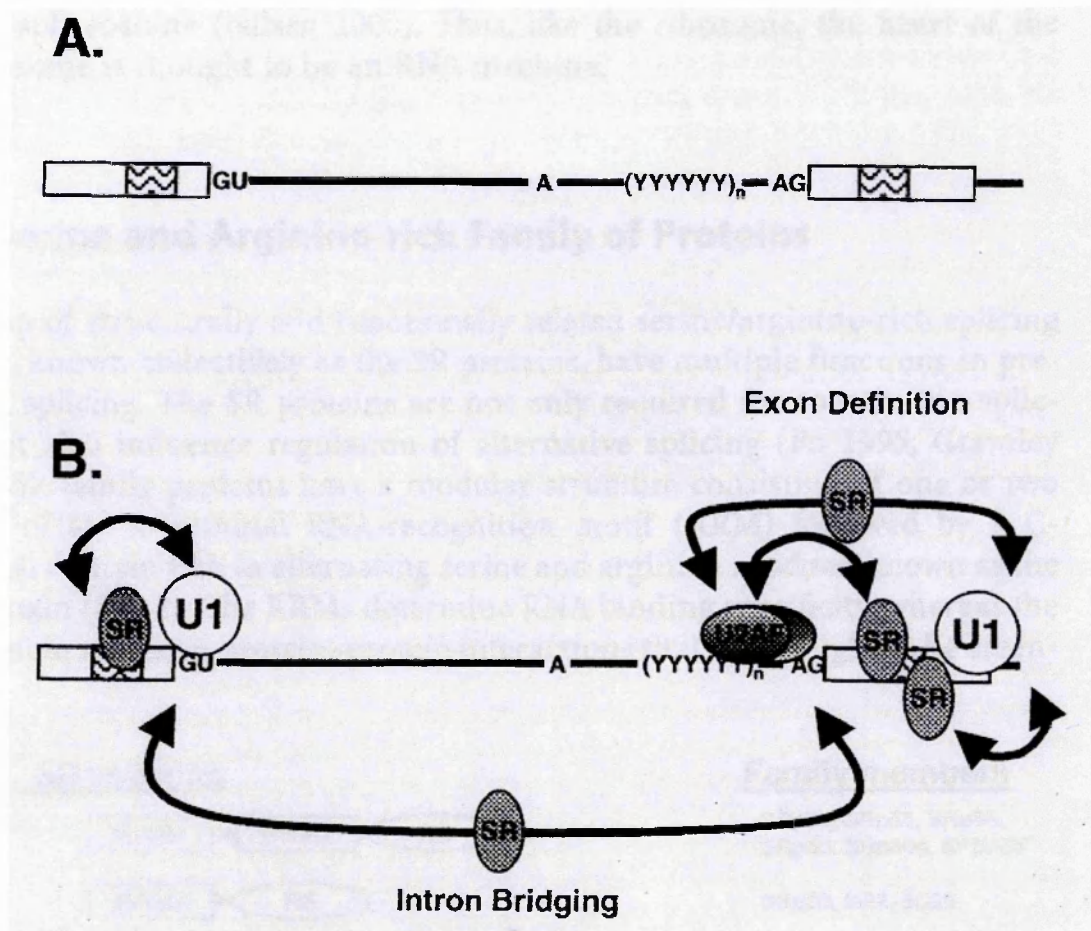


Figure 6: SR proteins are involved at numerous steps of pre-spliceosome assembly. (A) The organisation of splicing signals within a typical pre-mRNA. Exons (open boxes) may contain splicing enhancer elements (wavy boxes) that promote recognition of the nearby 5' or 3' splice sites (GU and AG, respectively). Additional signals consist of the polypyrimidine tract (Y) that resides between the branch point adenosine (A) and the 3' splice site. (B) The exon-dependent and -independent functions of SR proteins in pre-spliceosome assembly. SR proteins can promote both the recognition of 5' and 3' splice site as well as communication of splice sites by exon definition or intron bridging interactions. Arrows indicate RS domain-mediated interactions. (Taken from Sanford et al., 2003)

2001). One example that illustrates this mechanism is the HIV-1 tat exon3, which contains ESEs that are bound by SF2/ASF and SC35 and also contains an ESS that is recognised by hnRNPA1. Initial binding of hnRNPA1 to the ESS promotes further binding of hnRNPA1 upstream in the exon, preventing the binding of SC35 to the ESE, but not the binding of SF2/ASF. Therefore, the ESS suppressed SC35 but SF2/ASF-dependent splicing (Zhu et al., 2001). Besides promoting the use of alternative 3' splice sites, SR proteins can bind to exonic enhancers to promote the recognition of alternative 5' splice sites. Alternative splicing of the fruitless gene in *Drosophila* (*fru*) is regulated by Tra and Tra2 and depends on an exonic splicing enhancer (*fruRE*) consisting of three 13-nucleotide repeat elements, nearly identical to those that regulate alternative sex-specific 3'-splice site choice in the doublesex (*dsx*) gene (Lynch and Maniatis, 1996, Lam, 2003). The Tra, Tra2 and RBP1 (*Drosophila* homologue of SRp20) complex was directly implicated in the selection of this 5' splice site in *fruitless* pre-mRNA (Ryner et al., 1996), suggesting that SR proteins bound to upstream splicing enhancers can stimulate the binding of U1 snRNP to the downstream splice site through an interaction with U1-70K, an SR-related protein that is a stable component of U1 snRNP. In constitutive splicing, SR proteins are believed to participate in the definition of exons (Berget, 1995) by binding to the exon where they simultaneously interact with U2AF³⁵ at the 3' splice site and U1-70K in the 5' splice site. Some authors hypothesise that the majority of the constitutively spliced exons are recognised by this mechanism (Reed, 1996). In support of this model, SR protein binding sites are found in constitutive exons and some have been shown to act as constitutive splicing enhancers (Mayeda et al., 1999b; Schaal and Maniatis, 1999a). SR proteins can also negatively regulate the splicing of an intron by binding to an ESS. One example that illustrates this mechanism of regulation is a late mRNA in bovine papilloma virus 1 (BPV-1), that contains an array of two ESEs and an ESS between them. All three elements bind SR

proteins, and the SR proteins bound to the ESS are believed to sequester SR proteins bound to ESEs or to interfere with the bridging activities of the SR proteins bound to ESEs (Zheng et al., 1998). The context of the binding element is important in defining the activity of the SR protein bound to it. SF2/ASF inhibits adenovirus IIIa pre-mRNA splicing by binding to an intronic repressor element, but activates splicing of the same pre-mRNA when the repressor element is situated in an exon (Kanopka et al., 1996).

Exon-independent functions

The exon independent functions of SR proteins are those for which a requirement for binding to an exonic sequence has not been determined. For instance, SR proteins interacting with U1-70K and U2AF³⁵ act as a bridge between components bound to the 5' and 3' splice sites (Kohtz et al., 1994; Wu and Maniatis, 1993). SR proteins are also required for *trans*-splicing, in which the 5' and the 3' splice sites are contained in different molecules (Bruzik and Maniatis, 1995; Chiara and Reed, 1995). SR proteins also facilitate the recruitment of the U4/U6-U5 tri-snRNP to the pre-spliceosome (Cao et al., 1997; Fetzter et al., 1997; Tarn and Steitz, 1995; Teigelkamp et al., 1997). This recruitment may involve the interaction of SR proteins with SR-related proteins associated with the U4/U6-U5 tri-snRNP, such as the tri-snRNP associated proteins of 27kDa (Fetzter et al., 1997), of 65 and 110 kDa (Makarova et al., 2001), or the U5-100K associated protein (Teigelkamp et al., 1997). As mentioned earlier in this chapter (section 1.4.), SR proteins can restore splicing in extracts depleted from U1 snRNP (Crispino et al., 1994; Tarn and Steitz, 1994). The SR protein SF2/ASF can antagonise the effect of hnRNP A/B type proteins on the selection of alternative 5' splice sites in a concentration-dependent manner (Caceres et al., 1994; Mayeda et al., 1993; Mayeda and Krainer, 1992; Wang and Manley, 1995). SF2/ASF was

shown to do this by enhancing the binding of U1 snRNP to all the 5' splice sites, resulting in the selection of the 5' proximal splice site. On the contrary, hnRNP A1 binds co-operatively and indiscriminately to the pre-mRNA and interferes with the binding of U1 snRNP, resulting in the selection of the distal 5' splice site (Eperon et al., 1993; Eperon et al., 2000).

Other functions of SR proteins

When an mRNA is synthesized in the nucleus, it is immediately coated with a group of proteins, the hnRNPs. These proteins participate in a wide variety of processes in the nucleus (reviewed in Krecic and Swanson, 1999), and in the cytoplasm (reviewed in Shyu and Wilkinson, 2000). The cytoplasmic functions of hnRNPs include the regulation of cytoplasmic mRNA localisation, translation and mRNA turnover. It has been observed that, like the hnRNPs, a subset of SR family proteins including SF2/ASF, SRp20, 9G8, but not SC35 and SRp40, shuttle between that nucleus and the cytoplasm (Caceres et al., 1998). This observation suggested that, like hnRNPs, SR proteins could have functions in post-splicing mRNA processing, as well as functions in the cytoplasm. Recently, reports from the Steitz laboratory have confirmed this hypothesis. Using UV-crosslinking, transient transfections and *Xenopus* oocyte microinjection, it was shown that the shuttling proteins SRp20 and 9G8 interact specifically with a 22-nt RNA element from the histone H2a gene to promote the export of intronless RNAs in both mammalian cells and *Xenopus* oocytes (Huang and Steitz, 2001). More recently, these SR proteins and also SF2/ASF were shown to interact directly to the major receptor for bulk mRNA export to the cytoplasm, TAP/NXF1 (Huang et al., 2003), suggesting a role for SR proteins as adapters in mRNA

export. Probably in the years to come other roles of SR proteins in the cytoplasm will be reported.

1.10. SR-related proteins

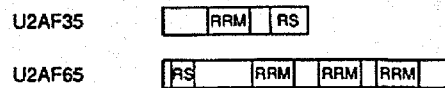
The SR-related proteins comprise a group of proteins containing an RS domain that have a distinct domain structure from the SR family of proteins (described in sections 1.5. and 1.8.). A genome-wide survey (Boucher et al., 2001) revealed a large complexity of RS domain-containing proteins in metazoans with functions not only in RNA processing, but also in chromatin structure, transcription by RNA pol II, cell cycle and cell structure (Boucher et al., 2001). This study also identified six RS-domain-containing proteins in *S. cerevisiae* with functions associated with cell structure, osmotic regulation and cell cycle progression. Interestingly, the vast majority of the RS related proteins identified are either known nuclear proteins, or related to nuclear-localised proteins. This is consistent with previous evidence indicating a role for RS domains in the nuclear-targeting of proteins (see section 1.7.). Most of the metazoan proteins analysed were linked to the transcription or processing of RNA pol II transcripts, but none of the proteins were known to be associated with the generation of either RNA pol I or pol III transcripts.

SR-related proteins and splicing

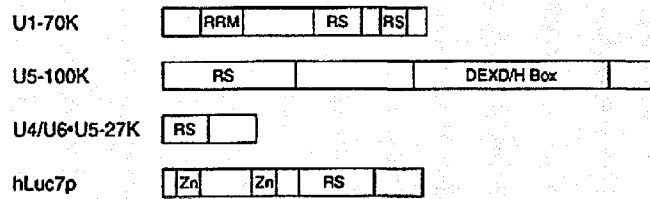
Most of the SR-related proteins with a known function are involved in pre-mRNA splicing (Fig. 7). Unlike SR proteins, the majority of them do not contain RRM domains and are not able to complement splicing deficient S100 extracts. One of the best characterised

B. Human SR Related Proteins

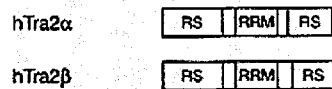
U2 Auxiliary Factor



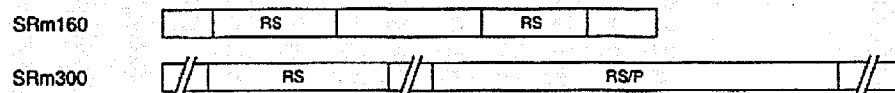
snRNP Components



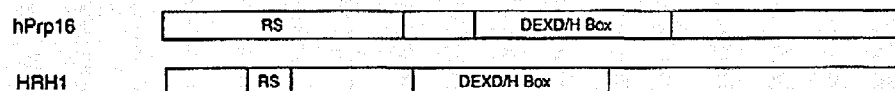
Splicing Regulators



Splicing Coactivators



RNA Helicases



Protein Kinases

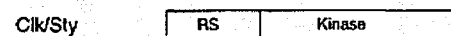


Figure 7: Schematic diagram of SR-related proteins. The domain structures for some of the human SR-related proteins that participate in pre-mRNA splicing are depicted. All proteins, with the exception of SRm300, are drawn to scale. RRM: RNA recognition motif; RS: arginine/serine-rich domain; Zn: zinc finger; DEXD/H Box: motif characteristic of RNA helicases. (Taken from Graveley, 2000)

examples are the subunits of the U2AF splicing factor, U2AF⁶⁵ and U2AF³⁵. The largest subunit, U2AF⁶⁵, contains three RRM domains responsible for binding to the polypyrimidine tract (Zamore et al., 1992) and an N-terminal RS domain, that facilitates the binding of the U2 snRNP to the branch point (Valcarcel et al., 1996). In contrast, U2AF³⁵, which contains an atypical N-terminal RRM followed by an RS domain, contacts the AG at the 3' splice site (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999) and stabilises the binding of U2AF⁶⁵ to the polypyrimidine tract (Zuo and Maniatis, 1996). An U2AF⁶⁵-like factor, HCC1, was identified using an autoantiserum from a hepatocellular carcinoma patient (Imai et al., 1993). HCC1 co-localises with other splicing factors in nuclear speckles and is very similar in size and structure to U2AF⁶⁵; however, its role in splicing remains to be established. A protein related to the splicing factor U2AF³⁵, Urp (for U2AF³⁵ related protein), was identified in a two-hybrid screen using an SR kinase (SRPK1) as a bait. Urp specifically interacts with U2AF⁶⁵ through a U2AF³⁵-homologous region and with SF2/ASF and SC35 through its RS domain. Therefore, Urp and U2AF³⁵ may independently position RS-domain-containing factors within spliceosomes (Tronchere et al., 1997). Other proteins related to U2AF³⁵ have been described. U2AF²⁶ shows a strong similarity to U2AF³⁵ but lacks an RS domain. Nevertheless, U2AF²⁶ can associate with U2AF⁶⁵ and functionally substitute for U2AF³⁵ in both constitutive and enhancer-dependent splicing (Shepard et al., 2002).

Mammalian SRm160/300 is a protein complex, consisting of two SR-related nuclear matrix proteins of 160 and 300 kDa, that functions as a coactivator of splicing (Eldridge et al., 1999). The SRm160/300 proteins both contain RS domains but lack recognizable RRM domains and do not normally bind to pre-mRNA in the absence of other splicing factors. The association of SRm160/300 with pre-mRNA requires U1 snRNP, SR family proteins, and is stabilized by U2 snRNP. Genetic evidence showed that the interactions between SRm160

and SR proteins are important for optimal splicing and proper development in *C. elegans* (Longman et al., 2001). SRm160 was identified as part of the exon-exon junction complex (EJC) although it is not a shuttling protein (Blencowe et al., 1998; Le Hir et al., 2000). The function of SRm160 in post-splicing events is not known. In contrast, RNSP1, an SR-related protein that act synergistically with SF2/ASF to stimulate both alternative and constitutive splicing *in vitro* (Mayeda et al., 1999a), is also involved in post-splicing events. RNSP1 was also identified as part of the EJC (Le Hir et al., 2000), and was demonstrated to be capable of triggering nonsense-mediated mRNA decay (NMD) when tethered downstream of a premature stop codon (Lykke-Andersen et al., 2001). Another SR-related protein that was shown to be essential for splicing is the SC35-interacting protein 1, Sip-1 (Zhang and Wu, 1998). This protein was originally identified by virtue of its interaction with SC35 in a yeast two hybrid assay, and subsequent assays indicated that this protein is also able to interact with other SR proteins and with U1-70K and U2AF⁶⁵.

The U1 snRNP-specific protein, U1-70K, harbors an RRM that is essential for binding to the U1 snRNA (Etzerodt et al., 1988; Mancebo et al., 1990; Query et al., 1989; Spritz et al., 1987; Theissen et al., 1986). It also contains an RS domain, which plays an important role in bridging 5' and 3' splice sites by interacting with SR proteins (Cao and Garcia-Blanco, 1998; Kohtz et al., 1994; Nelissen et al., 1994). The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that is required for spliceosome assembly and is phosphorylated by a yet unidentified snRNP-associated protein kinase (Fetzer et al., 1997).

CROP/Luc7A is an SR-related protein that was isolated as an overexpressed gene in a cisplatin-resistant cell line (Nishii et al., 2000). CROP/Luc7A contains leucine zipper-like motifs and a carboxy terminal RS domain. Its counterpart in yeast, Luc7p, was shown to be a component of the U1 snRNP and to be involved in 5' splice site selection (Fortes et al., 1999). Other examples of snRNP-associated proteins containing RS domains are the 65 and

110 kDa SR-related proteins of the U4/U6-U5 tri-snRNP, which were shown to be essential for the assembly of mature spliceosomes (Makarova et al., 2001), and the U5-100K protein (Teigelkamp et al., 1997). The U5-100K protein is homologous to *S. cerevisiae* Prp28p, a member of the DExD/H-box family of RNA-dependent ATPases that is required for conformational rearrangements in the spliceosome (Staley and Guthrie, 1999). Other SR-related proteins required for rearrangements of RNA-RNA and RNA-protein interactions in the spliceosome are the DEAH-box family proteins hPrp16 and HRH1 (Ohno and Shimura, 1996; Oortlepp et al., 1998), and the DEAD-box protein Hel117 (Sukegawa and Blobel, 1995).

Some of the kinases that phosphorylate SR proteins contain RS domain themselves. Clk1 has an RS domain at its N-terminus and interacts with several SR proteins and SR-related proteins in a yeast two-hybrid screen (Colwill et al., 1996b; Nestel et al., 1996). Clk1 was shown to regulate the splicing of *Clk1* and adenovirus pre-mRNAs *in vivo* (Duncan et al., 1997). Members of the Clk-Sty family of kinases were used to identify new SR-related proteins. Using Clk4 lacking kinase activity as a bait in a yeast-two hybrid screen, another SR-related protein, Clasp, was identified. This protein can be phosphorylated by Clk4, and promotes exon EB inclusion of Clk1 pre-mRNA *in vivo* (Katsu et al., 2002).

SRp30, SRp40 (SRp38) and SRp86 are members of the SR-related proteins family that act as splicing regulators antagonising SR proteins, and have been already described in this chapter (section 1.5.). SWAP, Tra and Tra2 were genetically defined as *Drosophila* alternative splicing regulators (reviewed in Fu, 1995). SWAP contains a C-terminal RS domain but lacks RRM, and autoregulates its expression by suppressing the splicing of its own pre-mRNA. Tra is an SR-related protein that lacks an RRM, whereas Tra2 contains a single RRM flanked by two RS domains. These proteins are alternative splicing factors that

function as part of a cascade of alternative splicing events involved in the regulation of sex determination in *Drosophila* (for a review see Schutt and Nothiger, 2000). The protein pinin, which contains an RS domain at the C-terminus as well as three coiled-coil domains in the first half of the protein, can regulate alternative 5' and 3' splicing (Wang et al., 2002). However, the mechanism by which pinin exerts this regulation remains unknown.

SR-related proteins and the transcriptional machinery

RNA processing events are coupled to transcription, and a key factor in this coupling is the carboxy terminal domain, or CTD, of RNA pol II (for reviews, see Bentley, 1999; Bentley, 2002; Maniatis and Reed, 2002; Proudfoot et al., 2002). Yeast two-hybrid screens using the CTD of RNA pol II as a bait resulted in the identification of a novel subgroup of SR-related proteins, collectively referred to as SCAFs (for SR-related CTD-associated factors) (Tanner et al., 1997; Yuryev et al., 1996). These proteins contain a short RS domain and a subset of them contain a single, divergent RRM. In addition, most of them contain a CTD-interaction domain (CID) that is 80-120 amino acids long. The mechanism by which this group of proteins link transcription to splicing remains unknown.

Several RNA pol II-associated factors, not previously recognised as harboring an RS, were identified in a genome wide survey of RS domain-containing proteins. These include a CTD phosphatase, proteins involved in transcriptional elongation and several factors involved in transcription initiation (Boucher et al., 2001). All these proteins are potential mediators in the communication between the transcription and splicing machines. It was demonstrated that the elongation rate of RNA pol II can influence alternative splicing (Kadener et al., 2001; Kadener et al., 2002; Nogues et al., 2002), and these proteins are good candidates to mediate this effect. Another surprising consequence of the coupling of

transcription and splicing is that the promoters that drive the transcription of a gene can influence the decisions made by the spliceosome hundreds or thousands of nucleotides downstream (Cramer et al., 1999). The molecular basis of this effect is unknown. One possibility is that the promoter somehow determines the elongation rate of RNA pol II. An alternative explanation is that factors that are specifically recruited to individual promoters can affect splicing directly. The later possibility is supported by the discovery of two SR-related proteins that bind promoters like classical transcription factors, PCG1 (Monsalve et al., 2000), which contains an RS domain and an RRM and can bind SR proteins directly *in vitro*, and the papillomavirus E2 protein (Lai et al., 1999). It is now known whether these proteins remain bound to the promoter and influence splicing at a distance or whether they are loaded onto the departing RNA pol II and influence splicing locally. Another protein that could be important in the communication between these machineries is NREBP/SON. NREBP/SON binds a negative regulatory element (NRE) located immediately upstream of the upstream regulatory sequence of core promoter and second enhancer of human hepatitis B virus (HBV), repressing the transcription of HBV genes and the production of HBV virions (Sun et al., 2001). NREBP/SON contains an RS domain and three groups on evolutionary conserved peptide repeats, and co-localises with the SR protein SC35 and with snRNAs in nuclear speckles (Wynn et al., 2000), although it is unknown if this protein has a role in splicing.

SR-related proteins and 3'-end processing

The processing of 3' end of pre-mRNAs is important for subsequent steps in the gene expression pathway including definition of terminal exons during splicing, mRNA stability, transport, and translation. It has been well documented that pre-mRNA splicing

and 3' end processing can influence each other. During splicing of the terminal 3' intron, splicing factors can interact with polyadenylation factors and promote 3' end cleavage and polyadenylation (McCracken et al., 2002; Vagner et al., 2000). Reciprocally, polyadenylation complexes are believed to promote splicing of the upstream 3' introns. This idea is supported by the observation that mutations of the poly (A) signal at the distal end of a 3'-terminal exon, inhibits splicing of proximal but not distal introns *in vitro* (Niwa and Berget, 1991). A good candidate for linking 3'-end formation and splicing is an SR-related protein, the 68 kDa subunit of the cleavage factor I (CFI) (Ruegsegger et al., 1998). CFI, together with CFII, the cleavage-polyadenylation specific factor (CPSF), the cleavage-stimulation factor (CstF) and the poly(A) polymerase, form a multiprotein complex that is responsible for executing the two steps of the polyadenylation reaction. Other factors that could link splicing and 3'- end formation via RS domain-mediated interactions are two SR-related proteins identified in *D. melanogaster* and *C. elegans* that are related to Fip1, a factor involved in the polyadenylation step of 3'-end formation in *S. cerevisiae* (Boucher et al., 2001).

SR-related proteins and cell structure

Several SR-related proteins are actin-binding proteins or are related to proteins with functions associated with the cytoskeleton. One of them, the nuclear matrix-associated protein SRm160/300, is discussed above. Another example is the Band 4.1 family of proteins. This family of proteins is believed to promote the anchoring of the cytoskeleton to the outer membrane by promoting the interaction of spectrin and actin in erythrocytes. Band 4.1 proteins localise to the nuclear speckles (Krauss et al., 1997) and interact with members of the SR family of proteins (Lallena et al., 1998). Another example of actin-

binding proteins that contain RS domains is the *S. cerevisiae* Sla1p protein (Boucher et al., 2001). It has been postulated that the function of actin or actin-like proteins in the nucleus is to form a nuclear homologue of the actin cytoskeleton that might facilitate the organisation of processes in the nucleus (Rando et al., 2000). Thus, SR-related proteins related to actin or actin-binding proteins could serve as anchors of the RNA processing machines and the nuclear matrix.

1.11. Identification of new factors involved in RNA processing

Although several components of the spliceosome and other RNA-processing machineries have already been characterised, many factors involved in these processes remain unidentified.

Historically, most mammalian RNA processing factors have been identified by genetics analysis, biochemical fractionation and purification, using antibodies recognising splicing factors, and by sequence homology. Developments in protein microcharacterisation using mass spectrometry have allowed the identification of many components of the spliceosome both in yeast (Gottschalk et al., 1998; Neubauer et al., 1997) and in mammals (Neubauer et al., 1998). This technology also allowed the identification of components of biochemically purified nuclear bodies, such as the interchromatin granules (IGC) which are components of the nuclear speckles (Mintz et al., 1999), and the nucleolus.

The spliceosome is one of the most studied complexes using these methods (Jurica and Moore, 2003). As described in section 1.4., spliceosome assembly proceeds through a series of sucomplexes or intermediates. The most common approach to date has been to

work with mixtures of assembly intermediates assembled *in vitro* (Gottschalk et al., 1998; Rappsilber et al., 2002; Zhou et al., 2002). The aim of this approach is to identify every factor involved, regardless at what stage of the splicing pathways it works. However, some recent reports described the purification and mass spec analysis of three spliceosomal subcomplexes (Hartmuth et al., 2002; Jurica et al., 2002; Makarov et al., 2002). This approach presents several drawbacks (Jurica and Moore, 2003). One of the problems is to determine which proteins are truly spliceosomal and which are contaminants. Some proteins can be detected in mass spec experiments simply because they are very abundant in the cell and contaminate any preparation. So, some proteins reported as spliceosomal factors could merely be contaminants, whereas some real spliceosomal components could have been mistaken as contaminants and therefore not included in the reports. Similarly, since splicing complexes contain RNA, some proteins reported could have been co-purified with spliceosomes simply because they bind RNA. Additionally, it is difficult to distinguish which proteins are specific pre-mRNA splicing factors and which are core factors, required for the splicing of every intron. Another problem is that if a protein is not detected by this approach, it does not necessarily mean that the protein is not a spliceosomal component. For instance, some of the U6 core snRNP proteins were undetectable in complexes that clearly contained U6 snRNA. Furthermore, many loosely associated factors can be lost during purification steps. Despite its well characterised role in the second step of splicing, Prp18 was undetectable in purified C complex by either mass spec or Western blotting (Jurica et al., 2002).

The gene trap methodology is an excellent experimental strategy that can be used to identify and mutate novel genes in order to uncover their biological function (for a review, see Cecconi and Meyer, 2000). Gene traps have been successfully used to isolate genes with interesting patterns of expression during development (Evans et al., 1997), or to obtain

sequence from large numbers of randomly trapped genes (Zambrowicz et al., 1998). This gene trap strategy can be modified in order to allow the identification of novel proteins of unknown function that localised to different subnuclear compartments, including splicing speckles (Tate et al., 1998).

1.12. Conclusions

In the preceding sections, some of the components that play a role in the biochemical reaction of splicing were described. Although numerous metazoan splicing factors have already been identified and characterised through biochemical and other approaches, it is clear that there is still a large number of factors that remain unidentified. The recognition and description of all the factors that take part of this process should help to fill the current gaps in the picture of splicing. The comprehensive analysis of the complete genomes of different organisms has uncovered many new RS domain-containing proteins with unknown functions. The functional characterisation of such proteins should lead to a dramatic increase in our knowledge of alternative splicing regulation and of the mechanism by which the spliceosome accomplishes the splicing reaction.

Materials and Methods

2.1. Plasmid constructs

All constructs used in this thesis were generated by PCR amplification of the fragment of interest and subsequent ligation into the corresponding vector. Useful restriction sites were incorporated in the primers used for PCR amplification with Vent (NEB) proofreading DNA Polymerase. PCR fragments were then purified using QIAQUICK PCR Purification Kit (QIAGEN) and digested with the corresponding restriction enzymes, which was carried out using the appropriate buffer and temperature recommended by the manufacturer (Roche). Ligation of DNA fragments was carried out in 10 µl reactions, containing approximately 100 ng of DNA (vector + insert) with a molar relation of 3:1 between insert and vector, and 400U of T4 DNA Ligase (NEB) and reactions were incubated o/n at 16°C. Normally, half of the ligation reactions were transformed into DH5α Subcloning Efficiency competent cells (Invitrogen), following manufacturer's protocol. Identity and orientation of the insert in positive clones were confirmed in every case by both digestion with restriction enzymes and sequencing (see section 2.3 for sequencing protocol).

2.1.1. Cloning and sequencing of mUSSRp58 and hUSSRp58

Mouse and human versions of USSRp58 (hereinafter, mUSSRp58 and hUSSRp58, respectively) were cloned by RT-PCR from total RNA of mouse ES cells or human Hep3B cells, respectively. Total RNA was prepared using Total RNA Isolation Reagent (ABgene)

according to manufacturer's specifications. Approximately 5 µg of total RNA were used for synthesis of first-strand cDNA with SuperScriptTMII RNase H⁻ reverse transcriptase (Invitrogen), following manufacturer's protocol, and 10% of the cDNA obtained in each case was used for PCR amplification. A fragment corresponding to the full-length coding sequence of mUSSRp58 was amplified using a forward primer 5'-TCGACTAGTGGACGCCGGTCTTCAGACACT-3' and a reverse primer 5'-CGAGGATCCTTAGGCCACGGGACTGCCCCAT-3', that introduce *SpeI* and *BamHI* restriction sites, respectively (sequence recognized by the restriction enzymes is underlined). For the human sequence, the following primers that also introduce *SpeI* and *BamHI* sites were used: Forward primer 5'-TCGACTAGTGGACGTCTGGTCATCAGATACT-3' and reverse primer 5'-CGAGGATCCTTAGGCCACAGGACTGCCCCAT-3'. These fragments were then digested with *SpeI* and *BamHI*, ligated to pBlueScript SKII⁻ and fully sequenced. These constructs were designated pBS-mUSSRp58 and pBS-hUSSRp58, respectively.

2.1.2. Epitope-tagged expression plasmids

The mammalian expression vector pCGT7 used to express mUSSRp58 and h-319RS has been previously described (Caceres et al., 1997). Transcription is driven by the CMV enhancer/promoter, and the coding sequence begins with an N-terminal epitope tag, MASMTGGQQMG, which corresponds to the first eleven residues of the bacteriophage T7 gene 10 capsid protein and is recognized by the T7-tag monoclonal antibody (Novagen). Due to the presence of an internal *XbaI* site in the USSRp58 cDNAs, the amplified fragments were designed with *SpeI* and *BamHI* sites and were subcloned into the *XbaI*-*BamHI* sites of pCGT7. The following primers were used:

mUSSRp58

Forward primer 5'-TCGACTAGTGGACGCCGGTCTTCAGACACT-3'

Reverse primer 5'-CGAGGATCCTTAGGACACGGGACTGCCCCAT-3'

hUSSRp58

Forward primer 5'-TCGACTAGTGGACGTCGGTCATCAGATACT-3'

Reverse primer 5'-CGAGGATCCTTAGGACACAGGACTGCCCCAT-3'

These plasmid constructs were designated pCGT7-mUSSRp58 and pCGT7-hUSSRp58, and were used for transient expression in a variety of cell lines.

A plasmid expressing hLuc7a was prepared in a similar way. A fragment encoding full-length hLuc7a was amplified by RT-PCR from total RNA purified from HeLa cells. The primers used to amplify hLuc7a were a forward primer 5'-GCTCTAGAATTTCGGCCGCGCAGTTGTTG-3' and a reverse primer 5'-CGGGATCCATTGGACTGAGTGTCACCTTC-3', that introduce *Xba*I and *Bam*HI sites, respectively. This fragment was digested with these restriction enzymes and ligated into the *Xba*I-*Bam*HI sites of pCGT7.

A fusion of mUSSRp58 to GFP was constructed by amplification of mUSSRp58 with specific primers and the resulting PCR product was subcloned as *Eco*RI-*Bam*HI fragment into either pEGFP-C1 (NH₂-terminal tag) or pEGFP-N1 (C-terminal tag)(Clontech). The fragment for pEGFP-C1 was amplified with the primer 5'-CGGAATTCGCCACGGGACTGCCCCAT-3' and the reverse primer 5'-CGGGATCCTTACGGCACGGGACTGCC-3', digested with *Eco*RI and *Bam*HI, and ligated. The primers used to amplify the fragment ligated to pEGFP-N1 were 5'-CGGAATTCAT**T**GGGACGCCGGTCTTCAGAC-3' as forward primer (the translation initiation codon is in bold). These two plasmid constructs were designated pEGFPN1-USSRp58 and pEGFPC1-USSRp58, respectively.

2.1.3. Plasmids used for yeast two-hybrid analysis

The bait plasmids used for yeast two-hybrid screening and direct yeast two-hybrid analysis were constructed by inserting PCR fragments into plasmid pGBKT7. The bait construct used for two-hybrid screening, designated as pGBKT7-mUSSRp58 was constructed by amplifying a USSRp58 cDNA with specific primers, and the resulting PCR product was subcloned as *EcoRI*-*BamHI* fragment into pGBKT7 (Clontech).

Primers:

Forward: 5'-TCGGAATTCGGACGCCGGTCTTCAGACACT-3'

Reverse: 5'-CGAGGATCCTTAGGCCACGGGACTGCCCCAT-3'

pGBKT7-319Nterm was constructed by amplification of mUSSRp58 (aminoacids 1-164) with specific primers and the resulting PCR product was subcloned as *EcoRI*-*SalI* fragment into pGBKT7 (Clontech).

Primers:

Forward: 5'-TCGGAATTCGGACGCCGGTCTTCAGACACT-3'

Reverse: 5'-ACGCGTCGACGCCTTTGATGCTGTGCACTT-3'

pGBKT7-319Cterm was constructed by amplification of mUSSRp58 (amino acids 165-334) with specific primers and the resulting PCR product was subcloned as *EcoRI*-*SalI* fragment into pGBKT7 (Clontech).

Primers:

Forward: 5'-CGGAATTCGGGGATTCTGGAAACATCAAA-3'

Reverse: 5'-ACGCGTCGACGGCCACGGGACTGCCCCA-3'

The constructs used for direct yeast two-hybrid analysis expressing SR proteins as preys were generated by inserting PCR fragments into plasmid pACT2 (Clontech). The following pairs of primers were used to amplify PCR fragments corresponding to the full-length coding sequence of different SR proteins:

SF2: 5'-TCCCCCGGGGTCGGGAGGTGGTGTGATT-3' and 5'-GGAATTCTTAGGTACGAGAACGGCAGCG-3', that introduce *Sma*I and *Eco*RI restriction sites respectively.

SRp55: 5'-C GGGATCCC GCGCGTCTACATA-3' and 5'-GGAATTCTTAATCTCTGGAAGTCTCGA-3', introducing *Bam*HI and *Eco*RI restriction sites, respectively.

SRp40: 5'-TCCCCCGGGGAGTGGCTGTCGGGTATTCATC-3' and 5'-GGAATTCTTAATTGCCACTGTCAACTGA-3', introducing *Sma*I and *Eco*RI restriction sites, respectively.

SRp20: 5'-CATGCCATGGATCGTGATTCCTGTCCATTG-3' and 5'-GGAATTCCTATTTCTTTTCATTTGACCTA-3', introducing *Nco*I and *Eco*RI restriction sites, respectively.

9G8: 5'-CATGCCATGGCGCGTTACGGGCGGTAC-3' and 5'-CGGGATCCTCAGTCCATTCTTTTCAGGACT-3', introducing *Nco*I and *Bam*HI restriction sites, respectively.

In each case, the PCR fragment was purified, digested with the appropriate restriction enzymes and ligated to the corresponding sites into the pACT2 vector.

2.1.4. Plasmids for expression in Baculovirus system

For overexpression and purification of recombinant m319RS in Baculovirus system, full-length mUSSRp58 fused to c-myc and 6xHis as carboxyl terminal tags was amplified by PCR and inserted into the plasmid pVL1392 (Pharming).

The following pairs of primers were used to amplify a PCR fragment: 5'-CGGAATTCATGGGTCGCTCTTCAGACACT-3' and reverse primer 5'-CGGGATCCC(ATG)₆GTCGACGGC-3'. These primers introduce *EcoRI* and *BamHI* restriction sites, respectively.

2.2. Antibodies

2.2.1. Production and purification of antibodies against USSRp58

An adult rabbit was injected with two synthetic peptides conjugated to KLH. Pre-immune bleeds were collected, and animals were injected with the mix of peptides conjugated to KLH, and boosted four weeks later. Two weeks after the boost the first bleed was collected. The animal was injected again two weeks after the first bleed, and the second immune bleeds were collected two weeks later. The same procedure of immunisation and bleed after two weeks was repeated two more times, until a satisfactory immunoresponse was obtained. The two peptides (Resgen) used for injection were the following:

- PEP1 (H-CEEEAKRRKEEDQATL-OH), corresponding to USSRp58 amino acid sequence 207 to 221.
- PEP2 (H-CLIALRQERLMGSPVA-OH), corresponding to USSRp58 amino acid sequence 319 to 334.

These peptides were resuspended in a solution containing 50 mM Tris pH 8.5, 5 mM EDTA and 10% glycerol, and two columns (one for each peptide) were prepared by coupling 2 mg of peptide to 1 ml of SulphoLink® Coupling Gel (Pierce), following manufacturer's protocol.

Approximately 20 ml of serum from the fourth bleed were used to affinity purify antibodies. Addition of caprylic acid to serum was used to precipitate most serum proteins with the exception of the IgG molecules. . Two volumes of 60mM NaAcO pH 4.0 were added, and the pH was adjusted to 4.8. Then, 1.5 ml of caprylic acid were added dropwise, and left to stir for 30 min. at RT. Then, the mix was centrifuged 10 min at 5000g and the supernatant was recovered and neutralized with 1/10th volume of 10X PBS. The supernatant was cooled to 4°C and 0.313g/ml of finely ground ammonium sulphate was added slowly while stirring for approximately 2 hours, to bring to a final concentration of 50% of saturation. Then the mix was centrifuged at 3000g for 30 min. After centrifugation, the supernatant was carefully decanted and the pellet (containing antibodies) was resuspended in 10 ml of PBS and dialysed overnight in 3 changes of PBS at 4°C. Next day, the dialysed IgG was loaded onto the column containing immobilised PEP1 (previously equilibrated with PBS), and the eluate was passed through the column two more times. Then, the eluate was applied to the second column, containing immobilised PEP2, and the same procedure applied on the first column was repeated with the eluates on this column. Columns were then washed with 20-bed volumes of 10mM Tris pH 7.5 and then with 20-bed volumes of 0.5 M NaCl, 10 mM Tris pH 7.5. Antibodies that were bound by acid-sensitive interactions were eluted by passing 10 bed-volumes of 100mM Glycine pH 2.5 through the column, passing 850 µl at a time, and collecting the fractions in an eppendorf tube containing 150 µl of 1 M Tris pH 8.0. The columns were then washed with 100 mM Tris pH 8.8 until the pH

of the eluate was 8.8. Then, antibodies that were bound by base-sensitive interactions were eluted in ten fractions of 850 μ l of freshly prepared 100 mM triethylamine pH 11.5. Fractions were collected in eppendorf tubes containing 150 μ l of 1 M Tris pH 8.0. Columns were then washed with 10 mM Tris pH 7.5 until the eluate reached a pH of 7.5. Columns were regenerated by washing with 8 bed-volumes of PBS, then equilibrated with deionized water containing 0.05% NaN_3 and 10 mM EDTA, and stored at 4°C. The protein content of the eluted fractions was monitored by reading the absorbance at 280 nm. The basic and acid fractions containing protein were pooled and dialysed overnight at 4°C against PBS plus 0.05% NaN_3 . The next day, the purified antibodies were adjusted to 10% glycerol, 1 μ g/ml BSA, 0.02% NaN_3 and stored at 4°C or frozen at -20°C for later use.

2.2.2. Other antibodies

Monoclonal antibody against T7-tag was obtained from Novagen. This antibody was used in a dilution 1:1000 for immunofluorescence experiments and in a dilution 1:10000 for Western blots. Monoclonal antibodies against SC35 (Fu and Maniatis, 1990), were kindly provided by Dr. Xiang-Dong Fu. These antibodies were used in a dilution 1:500 for immunofluorescence. Monoclonal antibodies (mAb96) against SF2/ASF (Hanamura et al., 1998) were provided by Dr. Adrian Krainer. These antibodies were used for Western blots in a dilution 1:500. The Y12 mAb (Lerner and Steitz, 1979) was kindly provided by Dr. Wendy Bickmore.

2.3. Sequencing of DNA

PCR fragments or plasmid DNA were sequenced using BigDye™ terminator cycle sequencing ready reaction (Applied Biosystems). A reaction usually contained approximately 500ng of DNA, 4 µl of reaction mix, 1 µl of primer 5 µM and dH₂O up to 10 µl. The sequencing program was run on a DNA engine PTC-200 (Peltier Thermal Cycler), and the program used consisted of an initial cycle of denaturation at 96°C for 3 min., followed by 25 cycles of denaturation at 96°C for 30 sec., annealing at 50°C for 15sec., and elongation at 60°C for 4min. The samples were then precipitated with EtOH, the pellet washed with 70% EtOH, and air-dried. The samples were then loaded on an ABI 377 machine, and the results analyzed using the ABI Sequencing Analysis 3.0 software.

2.4. Northern blot

To study the expression profile of 319RS in adult mouse tissues, a Multiple Tissue Northern (MTN®) was obtained from Clontech. According to the manufacturer, these blots are prepared with full-length poly A⁺ transcripts with virtually no genomic DNA contamination. The blots contain approximately 2 µg of purified poly (A)⁺ RNA per lane. The probe used to hybridise the blot was prepared by digesting pBS-mUSSRp58 with the restriction enzymes *Stu*I and *Xba*I, resulting in a fragment of 179 bp that would hybridise close to the 5' end of the coding sequence of USSRp58. This fragment was gel purified and labelled by random-priming using the HighPrime Kit (Roche), and hybridisation was performed following manufacturer's instructions.

For the detection of snRNAs, precipitated total RNA was separated in a 10% urea-PAGE, and transferred to Hybond N+ membrane (Amersham) using a Genie Blotter unit (Idea Scientific Company). The transfer was performed in 89 mM Tris-Borate, 2mM EDTA, pH 8.3, at 12 volts for 30 min. The immobilised RNA was then fixed to the membrane by incubation at 80°C for 2 hours. Small nuclear RNAs were detected with the following specific oligonucleotides:

U1 snRNA: 5'-ACCTTCGTGATCATGGTATCTCCCCTGCCAGGTAAGTATC-3'

U2 snRNA: 5'-GATACTACACTTGATCTTAGCCAAAAGGCCGAGAAGCGAT-3'

U4 snRNA: 5'-GATAAACCTCATTGGCACGATACTGCCACTGCGCAAAGCT-3'

U5 snRNA: 5'-AGGCGAAAGATTTATGCGATCTGAAGAGAAACCAGAGTAT-3'

U6 snRNA: 5'-ATCGTTCCAATTTTAGTATATGTGCTGCCGAAGCGAGCAC-3'

These oligos were labeled with polynucleotide kinase and purified using a Nucleotide Removal Kit (QIAGEN), following manufacturer's instructions. The blot was hybridised with a mix of the labelled probes in UltraHyb buffer (Ambion) at 42°C. Washes were performed as suggested by UltraHyb User's Manual, and the blot was exposed to an X-ray film.

2.5. Western blot

Samples were separated by SDS-PAGE and electroblotted onto Hybond-P (Amersham) in 25 mM Tris-base, 40 mM glycine and 20% methanol in a Genie Blotter unit (Idea Scientific Company), at 12 volts for one hour. The membranes were blocked with 5% nonfat dry milk in TBST (20 mM Tris pH 7.5, 137 mM NaCl and 0.1% Tween 20) for one hour at RT or overnight at 4°C. Incubations with primary and secondary antibodies were

carried out for one hour at RT in TBST containing 5% nonfat dry milk. Five washes with TBST were done after incubations with each antibody, and immunoreactive protein bands were detected with the SuperSignal system (Pierce) according to the manufacturer's instructions.

Protein samples from adult mouse tissues were prepared by homogenisation with T25 Basic Ultra Turrax (IKA Laboratechnik) in boiling lysis buffer (50 mM Tris pH 7.5, 10% glycerol, 0.05% SDS, 150 mM NaCl, 2.5% β -mercaptoethanol, 1 mM EDTA). Insoluble material was removed by centrifugation, and the protein concentration of each sample was determined using Bio-Rad Protein Assay system (Biorad) with BSA as a standard.

For preparation of protein samples from cells grown in culture, cells were washed with phosphate buffered saline (PBS) and collected by centrifugation. Cells were then resuspended in lysis buffer and boiled for 10 min and the protein content of the samples was determined as described above.

2.6. Cell culture and transfections

HeLa, 293T and 3T3 cell lines were grown in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal calf serum, and incubated at 37°C in the presence of 5% CO₂. HeLa and 293T cells were transfected with 3 μ g of DNA and 5 μ l of Lipofectamine 2000 (Invitrogen) per well in 6-well plates or with 15 μ g of DNA and 45 μ l of Lipofectamine 2000 per 10cm dish, according to manufacturer's instructions.

2.7. Indirect immunofluorescence

Cells were fixed and permeabilized for immunofluorescence assays at 24 hrs after transfection. Fixation was with 4% p-formaldehyde in PBS for 15-30 min at room temperature, followed by incubation for 10 min in 0.2% Triton X-100 in PBS. Fixed cells were then incubated for 1 hr at room temperature with the first antibody, washed with PBS, and incubated for 1 hr at room temperature with a secondary antibody, followed by further washes with PBS. The samples were observed on a Zeiss Axioskop microscope and the images were acquired with a Photometrics CH250 cooled CCD camera using Digital Scientific Smartcapture extensions in software from IP Lab Spectrum.

2.8. Nucleocytoplasmic shuttling assays

For transient transfections involving interspecies heterokaryons, due to the need for higher transfection efficiency, transfections were carried out by electroporation using 10 µg of plasmid DNA per 60-mm dish of 70-80% confluent cells in the presence of 20 µg of carrier DNA. Transfected HeLa cells were seeded on coverslips, followed by co-incubation with an excess of untransfected mouse NIH 3T3 cells for 3 hr in the presence of 50 µg/ml cycloheximide. The concentration of cycloheximide was then increased to 100 µg/ml, and the cells were incubated for an additional 30 min prior to fusion. Cell fusions were done as described (Pinol-Roma and Dreyfuss, 1992), and the heterokaryons were further incubated for 2 hr in media containing 100 µg/ml cycloheximide prior to fixation.

Immunofluorescence with the anti-T7 monoclonal antibody was performed as described above, except that DAPI was included at 5 µg/ml.

2.9. Yeast two-hybrid

The vectors and strains provided in the Matchmaker Two-Hybrid System 3 (Clontech) were used to screen a Pre-transformed Mouse Brain Matchmaker cDNA Library. The library is pre-transformed in *Saccharomyces cerevisiae* host strain Y187 (MAT α). The construct expressing the bait is transformed in the *Saccharomyces cerevisiae* reporter strain AH109 (MATa), which contains three reporters -*HIS3*, *ADE2*, and *MEL1* (or *LacZ*)- under the control of distinct GAL4 upstream activating sequences (UASs) and TATA boxes.

Manipulations of the yeast cells were carried out according to manufacturer's instructions (Clontech). In brief, the *Saccharomyces cerevisiae* reporter strain AH109 was transformed with the bait vector and subsequently tested for activation of reporter genes. Cells carrying the plasmid construct pGBKT7-m319RS were plated on either SD-Trp/-His, SD-Trp/-Ade, SD-Trp/X- α -Gal to check for activation of *HIS3*, *ADE2* or *MEL1* reporters, respectively. As the cells carrying the plasmid pGBKT7-mUSSRp58 showed some growth in the absence of histidine, the ability of this strain to grow in the presence on different concentrations of 3-aminotriazole (3-AT) was tested to determine the concentration of 3-AT at which the screen was performed.

2.9.1. Screen of cDNA libraries

The protocol used for screening the pre-transformed library is based on (Bendixen et al., 1994). A single colony carrying the pGBKT7-mUSSRp58 in AH109 strain was used to set up a 10 ml culture in Synthetic Dropout (SD) medium lacking Tryptophan (Trp), and the culture was grown overnight at 30°C with shaking. The next day, approximately 24 hrs after inoculation, this overnight grown culture was used to inoculate two 300 ml cultures of SD-Trp with 150 µl and 300 µl, respectively. The two 300 ml cultures were grown overnight at 30°C with shaking. The following morning, the culture with an OD₆₀₀ closest to 0.7 to 0.9 was chosen to continue with. A frozen aliquot of the library was resuspended directly with 1 ml of YPD medium and added to a flask containing 50 ml of YPD, and grown for 30 min at 30°C with shaking. The culture containing the library was then mixed with the culture of the cells containing the GAL4-BD Bait construct. Cells were then collected onto 0.22 µm GS sterile filters (Millipore) in screwtop bottle by syringing through the cells in 20 ml aliquots. Mating was then allowed to proceed by placing the filter discs, cell side up, on YPD-agar plates, and incubating at 30°C for 5 hrs. Cells were then resuspended from filters by placing them in a sterile petri dish and rinsing them up and down with 2 x 500µl of SD medium. All the washes were pooled in a tube, and 500 µl aliquots were plated in 14 cm plates containing SD -Trp/-Leu/ agar in the presence of 3 mM 3-AT, and incubated at 30°C for five days. Cells were also plated on SD -Trp, SD -Leu and SD -Trp/-Leu plates to calculate the mating efficiency and the number of library cells screened.

After 5 days, colonies growing in SD -Trp/-Leu/ agar in the presence of 3 mM 3-AT were picked and streaked out on plates containing SD -Trp/-Leu/-His/-Ade/X-α-Gal agar, and grown for another 5-8 days to check for expression of all three reporter genes (ADE2,

HIS3, and MEL1). The colonies that were Ade⁺, His⁺ and Mel1⁺ were considered positive interactors, and plasmid expressing the preys were then purified and sequenced. Phenotypes were confirmed by co-transformation of purified plasmids expressing the prey with either pGBKT7 or pGBKT7-mUSSRp58 into AH109. Interactors were considered true two-hybrid positives if they could activate the three ADE2, HIS3, and MEL1 reporters when co-transformed with pGBKT7-m319RS and were unable to activate them when co-transformed with pGBKT7.

2.9.2. Direct two-hybrid analysis

To directly test for the ability of two known proteins to interact, the pGBKT7 derived plasmids (expressing the protein of interest as fusion with GAL4 DNA-DB) and the pACT2 derived construct (expressing the protein of interest as a fusion with GAL4 AD) were co-transformed into AH109 strain. The resulting strains were tested for their ability to activate reporters, as described above.

2.10. Immunoprecipitation

For immunoprecipitation, 293T cells that were either transfected with a construct expressing the protein of interest or mock-transfected, were scraped from 10-cm tissue culture plates, collected in 1 ml of cold PBS, and centrifuged at 270g for 3 min in Eppendorf tubes. Cells were then resuspended in 800µl of IP buffer (50 mM Tris pH 7.5, 300 mM KCl, 5 mM EDTA, 0.5% Triton X-100, 1 mM DTT and 1x protease inhibitor cocktail), and incubated for 30 min at 4°C with continuous mixing on a rotary mixer. The

IP extract was centrifuged at 12000g for 20 min at 4°C, after which the pellet was discarded. The supernatant (IP extract) was divided into 200µl aliquots and either used immediately or frozen at -20°C for later use. To 200µl of IP extract, 1 µl (100 ng) of anti-T7 tag mAb (Novagen) or 10 µl of either immune or pre-immune serum and 20 µl of protein A or G (Amersham) were added. The IP extract was then incubated at 4°C for 2 hrs with continuous rotation. Immunoprecipitation reactions were then washed 4 times with the IP buffer and once with PBS. In some cases, after the last wash, beads were treated with 50 µg/ml of RNase A for 10 min at 4°C. After RNase A treatment, beads were resuspended in 40 µl of loading buffer (50 mM Tris pH 7.5, 10% glycerol, 0.05% SDS, 2.5% β-mercaptoethanol) and boiled for 10 min. For Western blot analysis of immunoprecipitated proteins, 10 µl of sample were used.

The protocol used for the analysis of co-immunoprecipitated RNAs has been described elsewhere (Makarova et al., 2001). Approximately 30 µl of affinity purified anti-USSRp58 antibody were bound to 20 µl of protein-A agarose, in the presence of 0.5 mg/ml of BSA, 50 µg/ml of total yeast tRNA, and 0.01% NP-40. Alternatively, PEP2 was added to the blocking solution at a concentration of 62.5 µg/ml. For immunoprecipitation of total snRNAs, 5 µl of Y12 mAb were bound to 20 µl of protein A agarose.

2.11. Detection of snRNAs

For analysis of co-immunoprecipitated RNAs, washed beads were phenol-chloroform extracted, and RNAs were precipitated with ethanol and 3' end-labeled with [³²P]pCp. Precipitated RNAs were incubated overnight at 16°C in a 10 µl reaction in the presence of 10 µCi of [³²P]pCp (Perkin-Elmer), 10U of T4 RNA Ligase (New England Biolabs) and 1x

Ligation Buffer, which was provided with the enzyme. The following day, labelled RNAs were precipitated, separated on 10% polyacrilamide-7 M urea gels and detected by autoradiography. Alternatively, U6 snRNA was detected by Northern blot, as described in section 2.4.

2.12. Purification of proteins

2.12.1. Purification of bulk SR proteins from Hela cells

Total SR proteins were purified from HeLa cells following a two salt precipitation protocol (Zahler et al., 1992). A pellet of approximately 10^9 HeLa cells (4C Biotech) was resuspended in 200 ml of Isolation buffer (10 mM Hepes pH 7.6, 65 mM KCl, 15 mM NaCl, 10 mM EDTA, 5 mM DTT, 5 mM KF, 5 mM β -glycerophosphate, 0.2 mM PMSF), and sonicated 10 times for 20 sec on ice. The suspension was centrifuged at 8000 rpm for 20 min at 4°C, transferred to new bottles and re-centrifuged. The supernatant volume was then measured and transferred to a new beaker. While stirring, ground ammonium sulphate was added to a final concentration of 65% saturation, and left to stir for two hours. The solution was then centrifuged at 8000 rpm for 20 min at 4°C. After centrifugation, the supernatant was transferred to a new beaker, and ground ammonium sulphate was added to a final concentration of 90% saturation and left to stir overnight at 4°C. The following day, the solution was centrifuged in SW28 rotor (Sorvail) at 25000 rpm for 1 hr. The pellet was then washed with isolation buffer containing 90% saturation of $(\text{NH}_4)_2\text{SO}_4$, resuspended in 10 ml of Dialysis buffer (the same composition of Isolation buffer, except with 1 mM EDTA and 1 mM DTT) and dialysed overnight against 2 changes of dialysis buffer. The

following day, the solution was aliquoted in Eppendorf tubes and centrifuged for 30 min at 13000 rpm at 4°C. The supernatants were transferred to new tubes and centrifuged for additional 20 min. The supernatants were then transferred to new tubes, MgCl₂ was added to a final concentration of 20 mM, and left on ice for 20 min. Tubes were centrifuged at 13000 rpm at 4°C, the pellet was washed with dialysis buffer containing 20 mM MgCl₂, and re-spinned for 15 min. Finally, the pellet was resuspended in 250 µl of BC100 (20 mM Tris pH 7.5, 100 mM KCl, 0.2 mM EDTA and 20% glycerol). The yield and purity of the preparation was determined by analysing 1 µl of the preparation on SDS-PAGE.

2.12.2. Purification of recombinant USSRp58

His-Myc-USSRp58 was purified from Sf9 baculovirus infected cells. The virus expressing His-Myc-319RS was generated using a BaculoGold Starter Package kit (Pharmingen), following protocols included in the User Manual provided with the kit. Three T-150 flasks containing Sf9 cells at approximately 80% of confluence were infected with 500 µl of a high titer stock of baculovirus expressing His-Myc-USSRp58. Four days after infection, cells were dislodged from the flasks, transferred to 50 ml conical tubes, and centrifuged at 450g for 5 min at RT. The supernatant containing the virus was discarded, and each pellet was resuspended in 10 ml of binding buffer (50 mM NaP buffer pH 8, 1.5 M NaCl, 5 mM Imidazole, 0.5% NP-40, 4 mM β-mercaptoethanol, and 1x of EDTA-free protease inhibitor cocktail). The suspensions of cells were then pooled and sonicated four times for 30 sec. on ice, followed by centrifugation at 17,000g for 20 min at 4°C. During the spin, 1 ml of slurry (=0.5 ml of resin) of Ni⁺²-NTA (Qiagen) was equilibrated by washing three times with 10 ml of binding buffer at RT. The cleared lysate was then incubated with the Ni⁺²-NTA beads for 1 hr at 4°C with gentle rocking, followed by a

centrifugation of 3 min at 450g at RT. The lysate was then removed, and beads were washed twice with 25 ml of washing buffer (50 mM NaP buffer pH 8, 1.5 M NaCl, and 20 mM Imidazole). Following washes, the beads were transferred to Eppendorf tubes, and the purified protein was eluted by applying 250 µl of elution buffer (50 mM NaP buffer pH 8, 1.5 M NaCl, 300 mM Imidazole), rocking 10 min at RT, centrifuging at 3,000 rpm for 1 min, and placing the eluate in a new tube. The elution step was repeated four more times. The eluted fractions were analysed for yield and purity on SDS-PAGE and the fractions containing protein were pooled. Imidazole and salts were dialysed away in a step-wise manner, by dialysing against BC800 first, followed by BC400 and BC100 (BC800 and BC400 have the same composition as in BC100, except they contain 800 mM and 400 mM KCl, respectively). Protein was then fractionated in 20 µl aliquots and stored at -80°C.

2.13. *In vivo* splicing assays

Transfection of Hela cells was done as described in section 2.6. Briefly, 1 µg of expression plasmid was co-transfected with 2 µg of the adenovirus E1A reporter plasmid pMTE1A (Zerler et al., 1986). Approximately 24hs after transfection, RNA was extracted using the Total RNA Isolation reagent (ABgene) following manufacturer's instructions. Total RNA was analyzed by RT-PCR with Superscript II reverse transcriptase (Invitrogen) and AmpliTaq DNA polymerase (Roche), as previously described (van Der Houven Van Oordt et al., 2000). E1A mRNA detection was carried out with the 5' end-labelled exon 1 forward primer 5'-GTTTTCTCCTCCGAGCCGCTCCGA-3' and the exon 2 reverse primer 5'-CTCAGGCTCAGGTTTCAGACACAGG-3'. Amplified products were separated by urea-PAGE, detected by autoradiography and quantified by PhosphoImage analysis.

2.14. *In vitro* splicing assays

2.14.1. Preparation of transcripts

Labelled transcripts were prepared in the presence of a CAP analog (m7G(5')ppp(5')G) (New England Biolabs) and [α - 32 P]GTP (PerkinElmer), as previously described (Sanford and Bruzik, 1999). Human β -globin substrate (Krainer et al., 1984) was linearised with *Bgl*II and transcribed with T7 RNA polymerase (Ambion). Fushi tarazu substrate (Rio, 1988) was linearized with *Xho*I and transcribed with T7 RNA polymerase. Transcripts were gel purified.

2.14.2. Immunodepletion of Hela nuclear extracts

In order to immunodeplete nuclear extracts of USSRp58, either anti-USSRp58 immune serum or pre-immune sera were crosslinked to protein G sepharose beads (Amersham). 200 μ l of beads were washed twice with PBS and then mixed with 200 μ l of either anti-319RS serum or pre-immune serum in 4 ml of PBS, and incubated for 1.5 hours at RT with gentle rocking. Beads were then washed twice with 5 ml of 0.2 M sodium borate. A sample of 5 μ l of beads were removed for later analysis, and the rest was resuspended in 5 ml sodium borate. Dimethylpimelimidate (DMP) was added to a final concentration of 20 mM, and incubated at RT for 30 min. After incubation, 5 μ l of beads were removed for later analysis. The reaction was then stopped by washing the beads with 0.2 M ethanolamine pH. 8, and then incubated in the same solution for 2 hours at RT. Beads were then washed once with PBS and incubated with 0.1 M glycine pH 3 at RT for 10 min. to remove all the uncrosslinked antibodies. The beads were then extensively

washed with PBS and stored in PBS + 0.01% NaN₃. The proportion of crosslinked antibody was monitored by analysing the samples isolated before and after the addition of DMP by SDS-PAGE and Coomassie staining.

The antibodies coupled to beads were washed five times with 1 ml of buffer D/1M KCl (20 mM Hepes pH 8, 20% glycerol, 0.2 mM EDTA, 1M KCl). HeLa nuclear extracts (4C Biotech, Belgium) were brought to 1 M KCl by directly dissolving the appropriate amount of crystals. The beads were then incubated for 2 hrs at 4 °C with 200 µl of nuclear extracts in Mobicolumns (MOBITEC), on a rotating wheel. The nuclear extracts were then removed by centrifugation, and kept on ice, while the beads were incubated at RT for 10 min with 100 mM glycine pH 3 with rocking. After 10 min, beads were washed 3 times with PBS and 3 times with buffer D (1M KCl), and two more rounds of depletion were applied to the nuclear extract. After three rounds of depletion were completed, the depleted nuclear extract was centrifuged at 14,000 rpm at 4°C for 5 min, and then was dialysed against two changes of buffer D+ (20 mM Hepes pH 8, 20% glycerol, 0.2 mM EDTA, 0.1 M KCl, 1mM DTT, 0.015% NP-40). After dialysis, the depleted nuclear extract was fractionated and stored at -80°C for later use. The protein concentration of the depleted extract was determined using Coomassie reagent (Biorad) with BSA as a standard. The extent of depletion was calculated by comparison of the depleted extracts with untreated extract in SDS-PAGE.

2.14.3. *In vitro* splicing

Splicing was carried out in 25 µl reactions containing 3.2 mM MgCl₂, 1 mM ATP, 20 mM creatine phosphate, 3% polyvinil alcohol, 30% nuclear extracts or 45% immunodepleted/mock-depleted nuclear extracts, complemented with buffer D with 0.1 M

KCl. Reactions were incubated at 30°C for 1 hr, and then RNAs were purified by proteinase K treatment, phenol-chloroform extraction and precipitation. Spliced products were analysed directly by electrophoresis on 5% denaturing polyacrilamide gels in Tris-Borate-EDTA (TBE) buffer. Gels were dried and exposed to PhosphorImager screens.

2.14.3. S100 complementation assays

Assays were performed in 25 µl reactions containing 5 mM MgCl₂, 5 mM ATP, 6.25 mM creatine phosphate, 3.65% PEG, 4.8% RNase inhibitors, and 50% S100 extracts (4C Biotech, Belgium), and were complemented with BC100 or His-Myc-USSRp58. Reactions were incubated at 30°C for 90 min. Spliced RNAs were purified and analysed as described above.

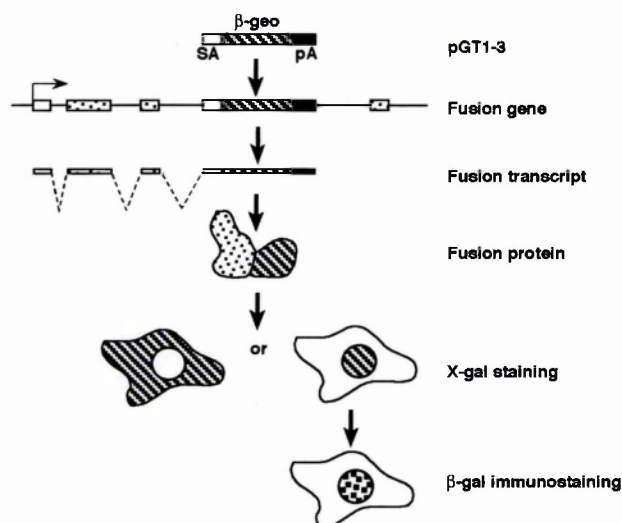
Results

1.1.Characterisation of a novel RS domain-containing protein

1.1.1. Identification and cloning of USSRp58

A gene trap screen in mammalian cells, in which the selection criterion is the subcellular localisation of the gene product, is being used in our laboratory to identify novel nuclear proteins. The gene trap system consist of a β -galactosidase-neomycin phosphotransferase (β geo) reporter gene, lacking its own promoter and ATG that when integrated into an intron of an expressed gene is spliced in frame into the gene transcript (Fig. 8A) (Tate et al., 1998). Drug selection identifies cells that have integrated the vector, in the correct orientation, into an intron of an expressed gene. This transcript encodes a fusion protein possessing both neomycin phosphotransferase and β -galactosidase activities (LacZ/ β -Gal), the latter of which can be detected by a simple histochemical stain (Fig. 8A). The value of this screen in mouse cells is that the initial cytological analysis with the light microscope gives immediate and detailed information about subnuclear localisation of the trapped gene. Integration of the gene trap β -geo construct into the mouse genome ensures that the tagged protein is expressed at endogenous levels and in the correct temporal fashion. A more detailed analysis of the subcellular localisation of the trapped fusion proteins is determined using antibodies against β -galactosidase. The sequence of the trapped gene is identified by 5' rapid amplification of cDNA ends (5' RACE) from the fusion transcript (Sutherland et al., 2001). The use of this approach allows

A



B

1	MGR RSSDTEEE SR SKRKKKHRR RS SSSSSSSD SR TY SR KKGGRR	43
44	PRSD SR SW SR DRQL RS HSYERRRRR RS SSSSSYG SR RK RS SR	86
87	SR GRGKPYRV QR SR SK SR TR RS SRPRPR RS HS SR SS ER SSHRRT	129
130	RS RS SR DRDRRKVRD KE KREREKDKGKD KE VHSIKRGDSGNKA	172
173	GLEHLPPAEQAKARLQLVLEAAAKADEALKAKER S EEEEAKRRK	215
216	EEDQATLVEQVKRVKEIEAIESDSFVQQTF RS SKDVKKAVEPS	258
259	EVQHVTAASGPASAAAEPPSTGKEIDPDSIPTAIKYQDDNSLA	301
302	HPNLFIEKAEAEKWFKRLIALRQERLMGSPVA	334

1	MGR RSSDTEEE SR SKRKKKHRR RS SSSSSSSD SR TY SR KKGGRR	43
44	PRSD SR SW SR DRQL RS HSYERRRRR RS SSSSSYG SR RK RS SR	86
87	SR GRGKPYRV QR SR SK SR TR RS SRPRPR RS HS SR SS ER SSHRRT	129
130	RS RS SR DRDRRKVRD KE KREREKDKGKD KE VHSIKR	164

Figure 8: Identification of a novel RS domain-containing protein. (A) Strategy for gene trap screen. pGT1-3 contain a β -geo cassette and neo^R enzymatic activities. This is preceded by a splice acceptor (SA) derived from the mouse *en2* gene and followed by a polyadenylation signal (pA). Productive integrations are those which occur in an intron of an expressed gene (fusion gene) downstream of coding exons (stippled boxes) such that the reporter is spliced in-frame into the gene transcript (fusion transcript). The sub-cellular distribution of the fusion protein is assessed by X-Gal staining. Sub-nuclear localisation of the fusion protein is determined by indirect immunofluorescence with anti-gal antibodies. **(B)** Amino acid sequence of mouse wild-type USSRp58 protein and the trapped protein (below), depicting the domains present. The RS domain is in white with a black background, whereas the small region rich in glutamic acid and lysine (EK-rich region) is in white with a gray background.

us to identify known proteins whose subnuclear localization was not previously recognized, and also to identify novel proteins located in a specific subnuclear compartment. One of the cell lines, number 319, generated in a screen performed in our laboratory by Kathryn Newton, expressed a fusion protein that localised to the nuclear speckles. I decided to study this gene, since localisation to the nuclear speckles is diagnostic of involvement in RNA processing. Using the sequence obtained by 5' RACE, an EST contig was generated that identified an open reading frame (ORF) predicting a putative protein. This sequence was used to design PCR primers that were used to amplify full-length cDNAs by RT-PCR from total RNAs of mouse ES cells and human Hep3B cells. These cDNAs were fully sequenced, and found to match the RIKEN cDNA clone 1200013F24 (sequence accession number gi: 27229009). The protein encoded by the gene was called USSRp58 for reasons described later in this chapter.

The protein contains 334 amino acids and includes an amino terminal domain rich in RS dipeptides and serine-rich regions, followed by a short glutamic acid-lysine (EK) rich region, similar to the one found in other splicing factors, such as the SR-related protein, SRrp86. The β -geo cassette was inserted in intron 4 of the USSRp58 gene resulting in a fusion protein that contains the N-terminal 164 aminoacids and conserves most of the RS dipeptides but misses the last 170 amino acids (Fig. 8B). This protein also possesses a bipartite nuclear targeting sequence between amino acids 214-230 that coincides with a helical region predicted to form coiled coil (180-237), a domain that often mediates protein-protein interactions. No homology to any other known protein motifs was found in the carboxy terminal half of the protein.

3.1.2. USSRp58 orthologs and paralogs in other species

The entire mouse and human genomes (MGSC3 and NCBI31 versions, respectively) were searched with the Rik1200013F4 sequence in order to find orthologous genes in humans and possible paralog genes in human and mouse. The human orthologous gene (Ensembl gene ENSG00000174891) mapped to human chromosome 3q25.32 (position ~159Mb in the NCBI31 assembly). There are no similar genes elsewhere in the genome, except for a highly similar sequence on chromosome 9q22.1 (position ~82.6Mb). This gene is described as an Ensembl gene (ENSG00000177872) and the predicted protein is 75% identical. However, a closer look showed that this gene has no introns, no start codon, and contains at least one in frame stop codon, all signs of a pseudogene. The same characteristics are present in the latest assembly (NCBI33). In mouse, there is also a second sequence hit in the genome, on chromosome 10 (position ~90.76 Mb). Again, the 5' and 3' end of the gene are missing (no start codon) and there are no introns, which again indicates that is most likely a pseudogene. In rat, there are 3 genes that are homologues to Rik1200013F4: Ensembl genes ENSRNOG00000012639, ENSRNOG00000018049, ENSRNOG00000004144. A bovine gene was also found in the TIGR database (TC132052).

In addition to these mammals, complete or partial gene sequences were detected in chicken, *Xenopus* (2 genes) and although zebrafish or *Fugu* orthologs were not found, a ortholog was found in medaka fish (*Orizias latipes* gene) and in trout (*Onchorynchus mykiss*). Also, one potential ortholog was found in the invertebrate chordate *Ciona intestinalis*. Several sequences showing certain homology to this gene were also found in *Drosophila melanogaster* and in *C. elegans*. An alignment of all the complete Rik proteins can be seen in Figure 9.

3.1.3. Production of recombinant USSRp58 and anti-USSRp58 antibodies

Despite several attempts I was unable to express this protein in *E. Coli* and only achieved very inefficient expression in the yeast *Pichia pastoris* (data not shown). Nevertheless, this protein could be expressed to satisfactory levels in baculovirus-infected Sf9 insect cells, and was purified from these cells as a histidine tagged protein over Ni-NTA agarose (Fig. 10). In addition to this histidine tag, a myc-tag was also introduced in the recombinant protein for use in future experiments.

Polyclonal antibodies were raised in rabbit against two synthetic peptides that are conserved between the human and mouse protein, encompassing amino acids 207 to 221 (²⁰⁷EEEAKRRKEEDQATL²²¹) and the last 15 amino acids of the protein (³¹⁹LIALRQERLMGSPVA³³⁴), which were designated PEP1 and PEP2, respectively. When the sequence of these peptides was blasted against either human or mouse translated genomes, the only matches obtained were for the genes that code for USSRp58, suggesting that they are likely to be unique to this gene. Serum from the immunised rabbit was affinity purified against each of the peptides coupled to a SulphoLink column (Pierce). After binding to the different columns, antibodies were recovered only from the column prepared with PEP2, suggesting that the animal used did not generate antibodies against PEP1.

The specificity of the anti-USSRp58 sera was assayed by Western blot analysis of extracts prepared from human 293T cells transfected with a construct expressing a T7 epitope-tagged version of wild-type human and mouse USSRp58. Western blot analysis of total cell extracts from transfected cells showed that a monoclonal antibody recognising the T7 epitope-tag revealed a single band of approximately 58 kDa for both the human and the mouse protein, which are identical at the amino acidic level (Fig. 11A). The deduced amino acid sequence predicts a protein of 334 amino acids with a calculated molecular mass of approximately 38.66 kDa and a isoelectric point of 11.9.

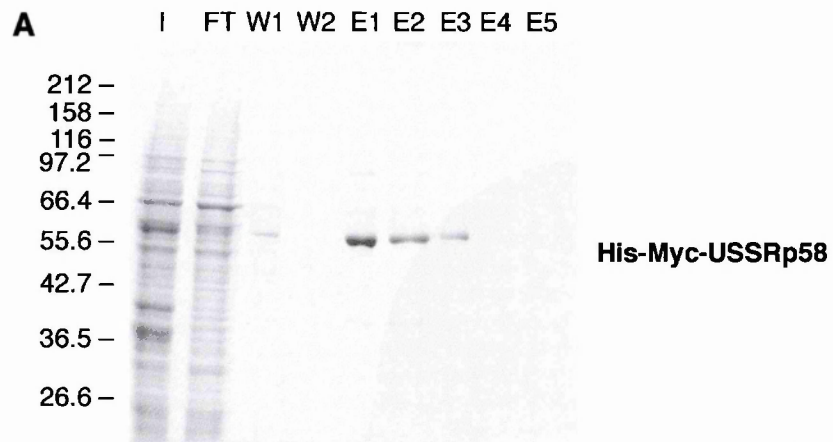


Figure 10: Purification of recombinant USSRp58 protein. His-Myc-tagged USSRp58 (His-Myc-USSRp58) was purified from baculovirus-infected Sf9 cells over Ni-NTA agarose. Samples were taken from the input (I), flow-trough (FT), washes (W1 and W2) and elutions (E1, E2, E3, E4 and E5), and the level of purification was determined by SDS-PAGE and Coomassie staining. Protein size markers (kDa) are indicated on the left.

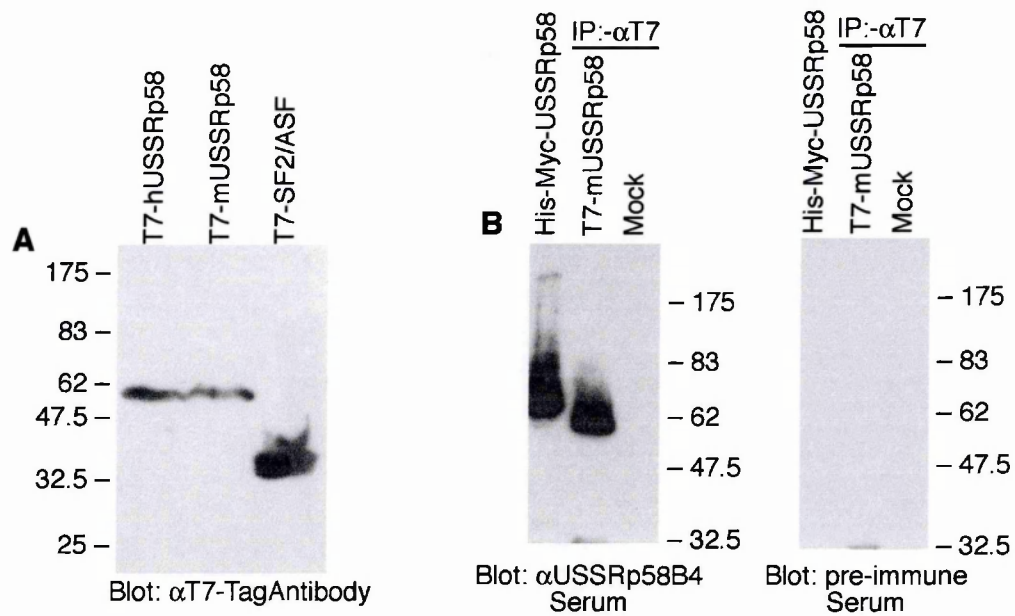


Figure 11: Characterisation of anti-USSRp58 antibodies. (A) 293T cells were transfected with plasmids expressing either T7-hUSSRp58, T7-mUSSRp58 or T7-SF2/ASF as a control. At 24 hrs post-transfection, whole cell extracts were prepared and analysed by Western blot using an anti-T7 monoclonal antibody (Novagen). Protein size markers (kDa) are indicated on the left. (B) Anti-USSRp58B4 (bleed 4) antibodies (left panel) or pre-immune serum (right panel) were used to probe blots containing baculovirus expressed His-Myc-USSRp58, and immunoprecipitates prepared with anti-T7 monoclonal antibody from extracts of 293T cells transfected either with T7-mUSSRp58 or mock transfected. Protein size markers (kDa) are indicated on the right.

Basic proteins have a high mass to charge ratio, resulting in a slower migration in SDS-PAGE. Presumably, potential phosphorylation of the serine residues in the RS domain of the protein also accounts for its anomalous migration at 58 kDa on SDS-PAGE. Similar disparities between predicted molecular mass and observed migration on SDS-PAGE have been reported for other RS domain-containing proteins, including SRp86, U1-70K, and SRp75 (Barnard and Patton, 2000; Chaudhary et al., 1991; Zahler et al., 1993).

The serum raised against the synthetic peptides (see above) was able to recognise the T7 epitope-tagged USSRp58 protein in total cell extracts of 293T transfected cells that were immunoprecipitated with the anti-T7 tag antibody. In contrast, no band was detected when used to probe an immunoprecipitate from mock-transfected cells (Fig. 11B). Furthermore, the serum was able to cross-react with the baculovirus expressed recombinant protein. When the same blot was probed with the pre-immune serum, no bands were revealed. Identical results were obtained with the affinity purified antibodies against PEP2 (data not shown). These results indicate that the antibodies raised against the synthetic peptides can specifically recognise the USSRp58 protein.

3.1.4. Subcellular localisation of USSRp58

The subcellular distribution of endogenous USSRp58 protein was analyzed by indirect immunofluorescence in HeLa cells. The affinity purified antibodies stained the nuclei of HeLa cells exclusively and showed a stronger signal in speckles (Fig. 12A). The signal observed with the affinity purified antibody was completely abolished when the antibodies were pre-absorbed with PEP2 (data not shown).

HeLa cells transfected with a construct expressing a GFP-tagged version of USSRp58, showed that the transiently expressed protein also localised exclusively in the

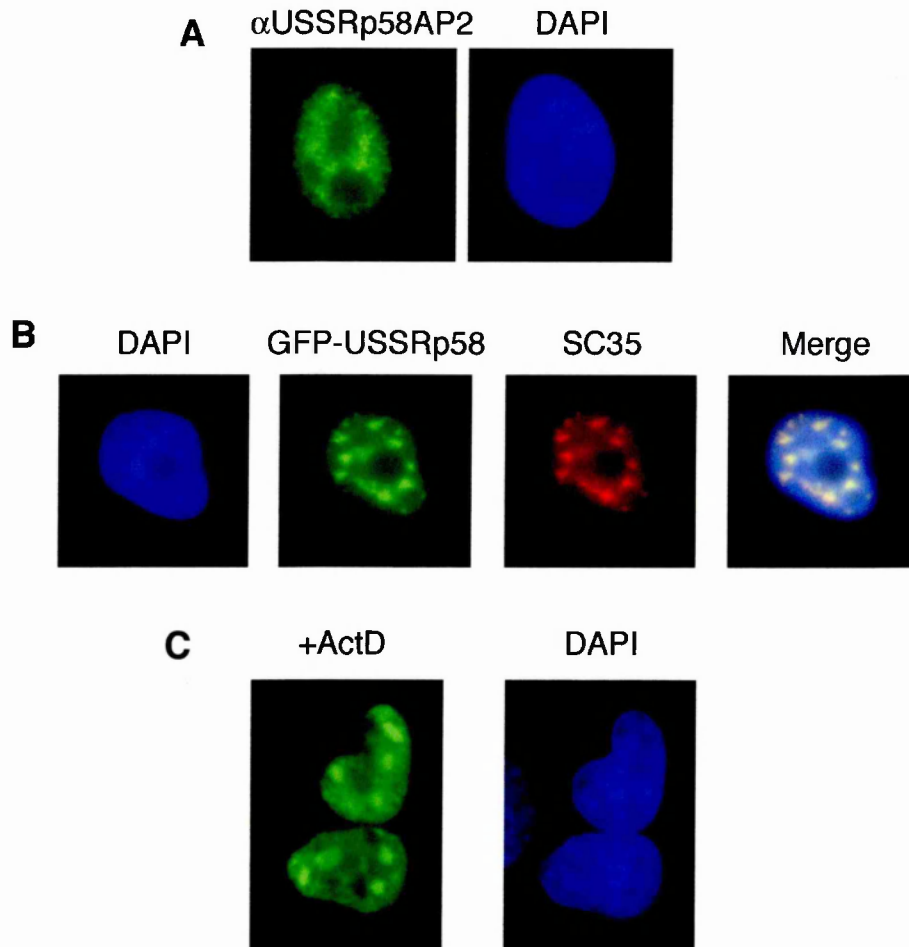


Figure 12: Subcellular localisation of USSRp58. (A) HeLa cells were fixed and the subcellular localization of endogenous hUSSRp58 was determined by indirect immunofluorescence with an anti-USSRp58 polyclonal antibody that was affinity-purified against PEP2 (anti-USSRp58AP2), followed by incubation with anti-rabbit IgG antibody conjugated with FITC. (B) HeLa cells were transfected with a plasmid encoding GFP-mUSSRp58, and fixed at 24 hrs post-transfection. Endogenous SC35 was detected by indirect immunofluorescence with an anti-SC35 monoclonal antibody and TxRD-conjugated secondary antibody (Fu and Maniatis, 1990). Images were superimposed to reveal sites of overlap in yellow (merge panel). (C) Effect of transcription inhibition on subcellular localization of USSRp58 protein. HeLa cells were transfected with a plasmid encoding T7-USSRp58-cells were incubated with actinomycin D (+ActD) plus cycloheximide for 3 hrs and fixed at 24 hrs post-transfection. The localisation of the expressed proteins was determined by indirect immunofluorescence with anti-T7 monoclonal antibody and FITC-conjugated secondary antibody.

nucleus, with a typical speckled pattern (Fig. 12B). Furthermore, GFP-USSRp58 colocalises with an SR protein, SC35, in a typical speckled pattern (Fig. 12B). Identical results were obtained when HeLa cells were transfected with a construct expressing a T7 epitope-tagged version of wild-type USSRp58 (data not shown).

Speckles are dynamic structures, from which splicing factors are thought to be recruited to sites of active transcription (Misteli et al., 1997). Arrest of transcription by RNA polymerase II inhibitors such as actinomycin D, cause speckles to become enlarged and rounded up (Spector et al., 1993). Indeed, this was observed upon addition of 5µg/ml Actinomycin D for 3h to HeLa cells transfected with a T7-tagged version of wild type USSRp58 USSRp58 (Fig. 12C). Taken together, these results suggest that USSRp58 may be involved in RNA processing and most likely in pre-mRNA splicing.

3.1.5. USSRp58 shuttles between the nucleus and the cytoplasm

It has been shown that many nuclear proteins involved in splicing, as for example several hnRNPs (Bennett et al., 1992; Pinol-Roma and Dreyfuss, 1991) and a subset of SR proteins (Caceres et al., 1998), shuttle between the nucleus and the cytoplasm. This fact suggests that in addition to their nuclear roles in splicing, shuttling proteins may have additional post-splicing functions. Many hnRNPs were shown to regulate at least three distinct cytoplasmic events: mRNA localisation, mRNA turnover and translation (Gama-Carvalho and Carmo-Fonseca, 2001; Shyu and Wilkinson, 2000). Members of the SR protein family were also shown to have roles in post-splicing RNA processing as well. Two shuttling SR proteins, SRp20 and 9G8, promote mRNA export (Huang and Steitz, 2001), most likely acting as adapter proteins for TAP-dependent mRNA export (Huang et al., 2003).

Many shuttling proteins accumulate in the cytoplasm when cells are incubated with inhibitors of RNA pol II transcription, such as actinomycin D. This effect relies on the fact that nuclear re-import of the shuttling protein requires ongoing transcription, and this property has been used to identify shuttling proteins (Dreyfuss et al., 1993). The localisation of several nucleo-cytoplasmic splicing factors, such as SF2/ASF, SRp20 and hnRNP A1, changes from the nucleus to the cytoplasm after treatment with actinomycin D, whereas others, such as 9G8, remain nuclear, despite the fact that are indeed shuttling proteins (Caceres et al., 1998). Incubation of transiently transfected HeLa cells with actinomycin D and cycloheximide revealed that USSRp58 remained nuclear under these conditions. This experiment alone does not rule out the possibility that USSRp58 shuttles from the nucleus to the cytoplasm, but that its re-import is independent of transcription.

To determine whether USSRp58 has the ability to shuttle between the nucleus and the cytoplasm, I used interspecies heterokaryons assays (Caceres et al., 1998; Michael et al., 1995) (Fig. 13A). HeLa cells were transfected with epitope-tagged USSRp58 and were then fused to mouse NIH 3T3 cells to form heterokaryons. Prior to fusion, the cells were treated with cycloheximide, so that no further protein synthesis takes place in the heterokaryons. At 2hr post-fusion, the cells were fixed and stained to examine the distribution of the tagged proteins by immunofluorescence microscopy with a monoclonal antibody directed against the epitope tag. To distinguish the human and mouse nuclei, the cells were stained with the dye DAPI, which gives a characteristic staining of heterochromatin in the mouse nuclei. Detection of the tagged protein – originally expressed in the human cells – within the mouse nuclei in the heterokaryons is indicative of shuttling (and therefore of nuclear export). Analysis of heterokaryons formed between human cells expressing T7-tagged USSRp58 and mouse acceptor cells, clearly showed that epitope-tagged USSRp58 was detectable in both nuclei (Fig. 13B). These data indicate that USSRp58 shuttles

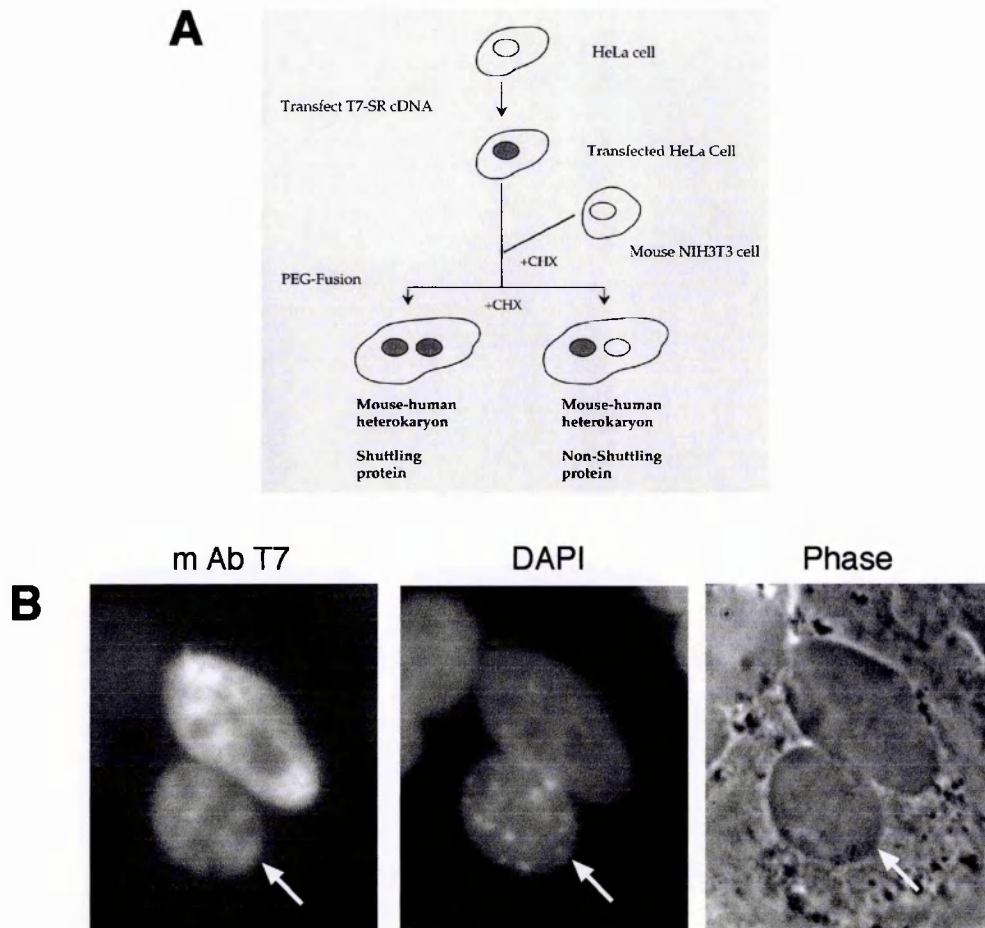


Figure 13: Analysis of nucleocytoplasmic shuttling of USSRp58 by transient transfection in interspecies heterokaryons. (A) Diagram of the experimental approach. The dark nuclei indicate localisation of the transiently expressed human SR protein. Cycloheximide (CHX) was added prior to fusion to prevent further protein synthesis from the human mRNA in the heterokaryons. **(B)** Detection of the transiently expressed USSRp58 protein in interspecies heterokaryons. HeLa cells were transfected with pCGT7-mUSSRp58. At 24 hr post-transfection, the cells were treated with cycloheximide and subsequently fused with mouse NIH 3T3 cells in the presence of polyethylene glycol, to form heterokaryons. The cells were further incubated for 2 hr in the presence of cycloheximide, followed by fixation. The localization of the expressed proteins was determined by indirect immunofluorescence with anti-T7 monoclonal antibody and FITC-conjugated secondary antibody (left panel). The cells were simultaneously incubated with DAPI for differential staining of human and mouse nuclei within heterokaryons (center panel). The arrows indicate the mouse nuclei within human-mouse heterokaryons. Phase-contrast images of the same heterokaryons are shown (right panel).

between the nucleus and the cytoplasm, suggesting that this protein could have a role in post-splicing mRNA processing.

3.1.6. Analysis of the expression of USSRp58

The expression of many splicing factors is regulated in a tissue or developmental specific manner. For example, the level of expression of SF2/ASF and that of its antagonist, hnRNP A1, vary naturally over a wide range in tissues and in immortal and transformed cells lines, supporting the notion that changes in the ratio of regulators of splicing can affect alternative splicing of different pre-mRNAs *in vivo* (Hanamura et al., 1998). Northern blot experiments were performed to determine the expression of USSRp58 at the RNA level, using a partial fragment of USSRp58 cDNA as a probe. This experiment showed that USSRp58 is widely yet differentially expressed across a variety of adult mouse tissues (Fig. 14). Interestingly, three different mRNA isoforms were detected. The major isoform migrates at about 3.2 kbp and is strongly expressed in testis, liver and brain, and weakly expressed in heart, spleen lung, kidney and skeletal muscle. This isoform most likely corresponds to cDNA RIKEN NM_025822 (Accession number gi:27229009), which encodes full-length USSRp58. The two other isoforms are strongly expressed in testis, and are also detected in liver and heart; however, no expression could be detected in skeletal muscle. One of these two isoforms could correspond to cDNA RIKEN AK005242 (Accession number gi:12837659), since it should be detected by the probe used. This cDNA is predicted to express a shorter isoform of USSRp58 of 222 amino acids, comprising the RS domain and the short EK-rich domain. The identity of the third band detected is unknown. The integrity of the RNA from each tissue was verified by reprobing

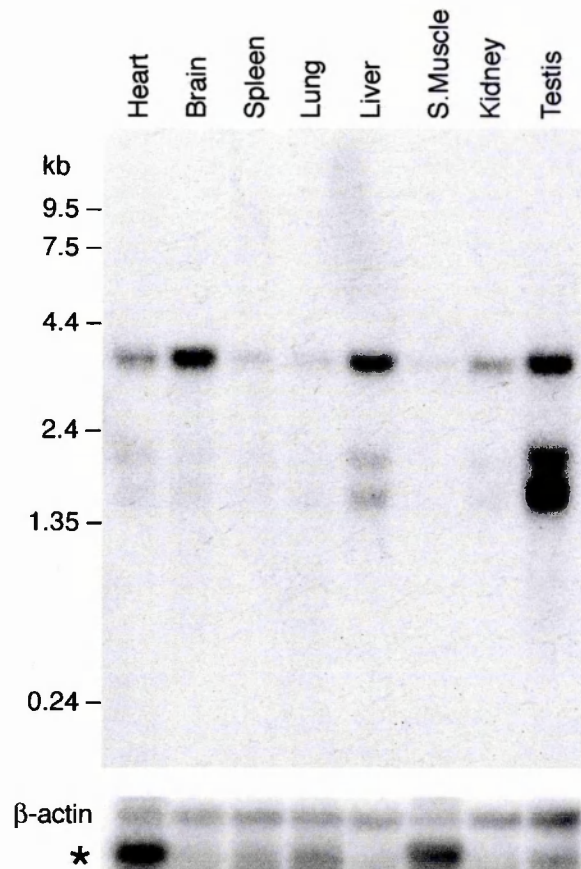


Figure 14: Northern blot analysis of mouse USSRp58. A Multiple Tissue Northern blot (Clontech) was probed with a ^{32}P -labelled probe prepared from mUSSRp58 cDNA and exposed to autoradiography (upper panel). The size of RNA markers is indicated on the left. beta-actin cDNA was used as an internal control (lower panel). Note that the human beta-actin cDNA probe strongly cross hybridises to muscle-type actin (marked with an asterisk) in heart and skeletal muscle.

the same filter with a human β -actin probe, that strongly cross-hybridises with mouse β -actin mRNA.

The predicted molecular mass of USSRp58 is 38 kDa, but Western blot analysis with anti-USSRp58 peptide antibodies (described above) detected a band of approximately 58 kDa in total cell extracts from human and mouse cells, suggesting that USSRp58 may be postrationally modified (Fig. 15A). When extracts prepared from different adult mouse tissues were analysed in the same way, two major isoforms were observed in all the tissues analysed (Fig. 15B). The lower band revealed by the antibody corresponds to USSRp58, according to the apparent size of the band. A second uncharacterised band of approximately 66 kDa cross-reacts with anti-USSRp58 antibody. This second band could be also observed in HeLa nuclear extracts (see Figure 21A). To test if this band corresponds to an hyperphosphorylated version of USSRp58, HeLa nuclear extracts were treated with calf intestinal phosphatase (CIP) and analysed by Western blot. Surprisingly, after treatment with CIP, both bands are not further detected by anti-USSRp58 antibody (data not shown). The reasons for this result are not known at the moment. Possibly, anti-USSRp58 antibody could be recognising a phosphoepitope present in USSRp58. Other explanation could be that after treatment with CIP, the overall charge of USSRp58 changes considerably, changing the migration properties of USSRp58 in the transfer buffer. This hypothesis will be tested in the near future. The 66 kDa cross-reacting band could also correspond to an alternatively spliced isoform of USSRp58, or alternatively represent a post-translational modification other than phosphorylation. As mentioned in section 3.1.3., the sequences of the peptides used to raise the antibodies were used in a BLAST search against the human and mouse translated genomes. The only match obtained in each case was the gene that expresses USSRp58, suggesting that is unlikely that anti-USSRp58 antibody would recognise a completely unrelated protein, although this possibility cannot

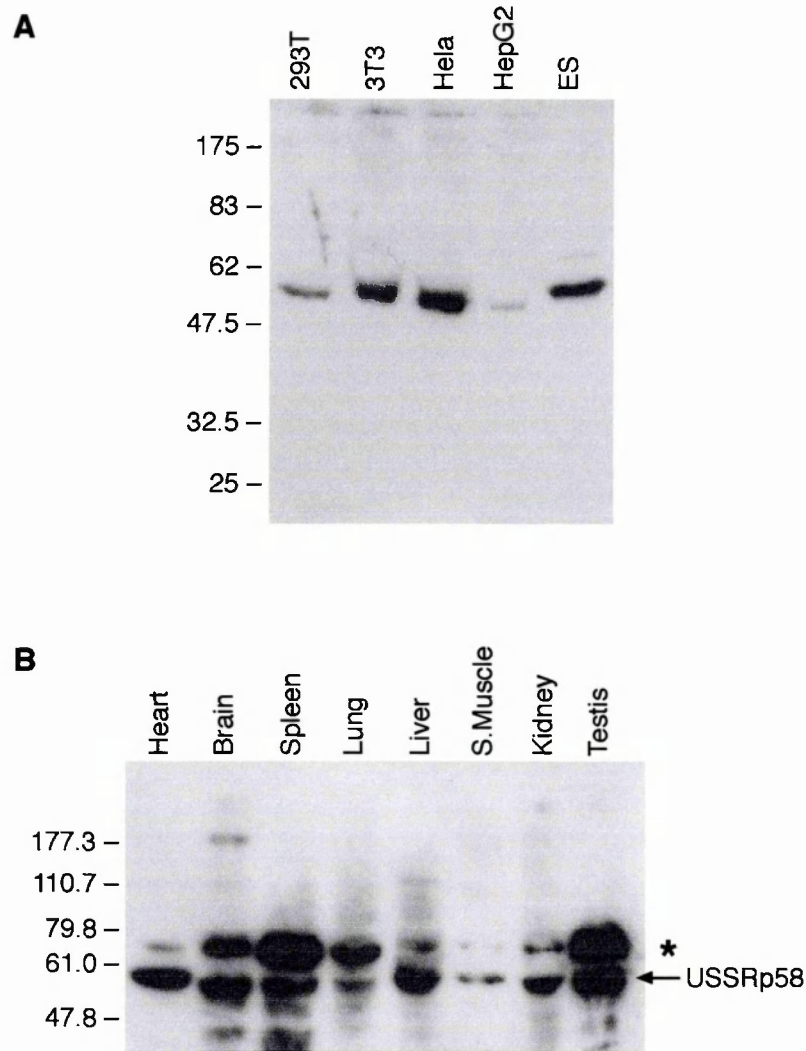


Figure 15: Expression of USSRp58 in tissues and cultured cells. (A) 10 ug of total protein from different cell lines were separated by SDS-PAGE and transferred to a nitrocellulose membrane. The membrane was then probed with anti-USSRp58B4 (Bleed 4) antibody. Protein size markers (kDa) are indicated on the left. **(B)** 50 ug of total protein from each of the mouse tissues indicated above were analysed as described in (A). The band corresponding to the uncharacterised protein recognised by anti-USSRp58B4 antibody is marked with an asterisk. Protein size markers (kDa) are indicated on the left.

be discarded at present. This band is currently being characterised by different means. The results obtained from Northern and Western blot analyses showed that USSRp58 is ubiquitously expressed but shows a tissue-specific or cell type-specific level of expression, as it has been reported for many other splicing factors (Cowper et al., 2001; Hanamura et al., 1998; Screaton et al., 1995; Shepard et al., 2002). This is consistent with a potential role for USSRp58 as a splicing regulator.

3.1.7. USSRp58 interacts with splicing factors in two-hybrid analysis

In order to identify interacting proteins, yeast two-hybrid screens were carried out with the full length sequence of USSRp58 as a bait (Chien et al., 1991). The cDNA encoding the full-length mouse USSRp58 protein was cloned into the pGBKT7 vector (Clontech), resulting in the pGBKT7-USSRp58 vector. This construct, that would express full-length USSRp58 as a fusion protein with the DNA binding domain of GAL4 transcription factor, was transformed in the *Saccharomyces cerevisiae* strain AH109, and used to screen a mouse brain cDNA library pretransformed into the *S. cerevisiae* strain Y187 (MATCHMAKER System 3, Clontech). At least 1.5×10^7 individual library plasmids were screened. The system provides three interaction reporters that can be tested. Positive clones that could activate the three reporters were grown, plasmids were isolated, and the interaction was confirmed by re-transformation of the purified plasmids. Once the interaction was confirmed, library plasmids were sequenced and the identity of the interacting protein was determined.

In support for a role of USSRp58 in pre-mRNA splicing, I identified Luc7p and HCC1, two splicing regulators, as interacting proteins. Most of the clones obtained are derived from these two genes and other two RS domain-containing proteins (Fig. 16A).

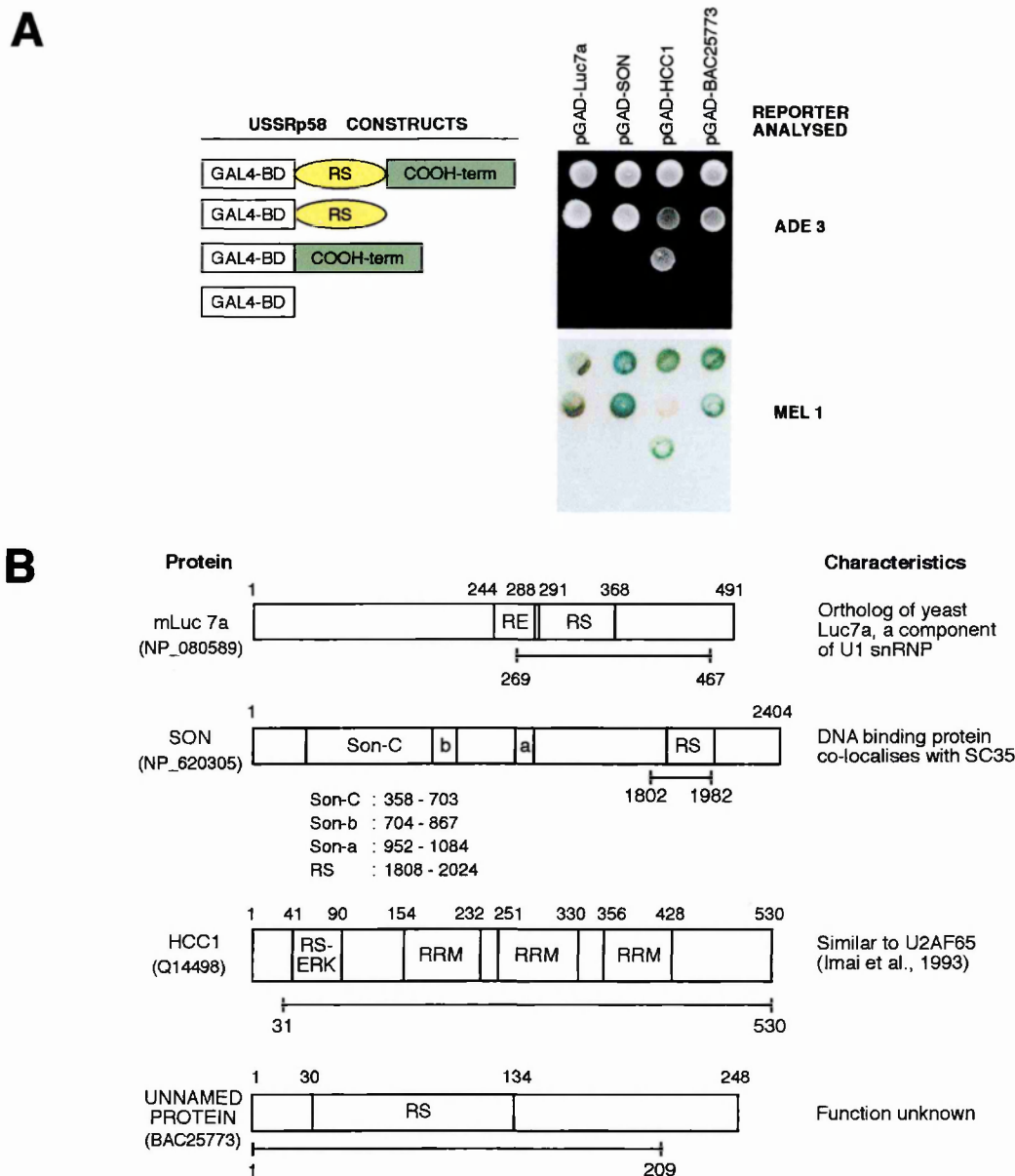


Figure 16: USSRp58 interacts with RS domain-containing proteins in a yeast two-hybrid system. (A) Reconstitution of two-hybrid interactions found in a yeast two-hybrid screen. The upper panel shows growth on selective media without Adenine, and the lower panel shows activation of the *MEL1* reporter. In each panel, yeast cells were co-transformed with each of the prey plasmids (pGAD) isolated from the screen (indicated above) and bait vectors expressing either full-length USSRp58 (first row), the N-terminal half of the protein comprising the RS domain (second row), the C-terminal half of the protein (third row), or the empty vector (fourth row). **(B)** A cartoon illustrates the primary structure, main features and amino acid length of the interactors found in the two-hybrid screen. The portions of each protein covered by selected clones are indicated by bars (with the N- and C- terminal amino acid numbers shown). The known features of the proteins are described on the right.

The Luc7 gene was identified in yeast in a synthetic lethal screen with the nuclear cap binding complex (CBC), and was shown to be a component of the U1 snRNP and to be involved in 5' site selection (Fortes et al., 1999). The Luc7 protein contains arginine-serine and arginine-glutamic acid repeat sequences characteristic of SR and SR-related proteins and two zinc finger motifs. Luc7 seems to have undergone duplications several times during evolution, and three Luc7 genes (Luc7a-c) are present in the human genome. Interestingly, human Luc7 does not seem to be a stable U1snRNP component in humans. CROP/Luc7a was isolated as an overexpressed gene in a cisplatin-resistant cell line (Nishii et al., 2000). CROP/Luc7a contains a leucine zipper-like motif and a carboxy terminal RS domain. HCC1, a 64 kDa nuclear autoantigen that harbors splicing factor motifs, was originally identified using an autoantiserum from a hepatocellular carcinoma patient (Imai et al., 1993). HCC1 localizes to the nuclear speckles and contains an N-terminal RS domain, and three RRM. The modular structure of HCC1 and the presence of an octapeptide sequence, the RS-ERK motif, within its RS domain is reminiscent of the large subunit of the U2AF splicing factor, U2AF⁶⁵. Another interacting protein, SON, contains a domain rich in RS dipeptides and localises to nuclear speckles. This protein also contains a series of domains defined by reiterated peptide repeats, which are unique to this protein and are not found in any other proteins in the databases. However, its function in RNA processing is presently unknown (Wynn et al., 2000). Finally, a fourth factor containing an RS domain encoded by a novel cDNA, Rik270043I21, was isolated. The function of this factor is completely unknown.

All the identified interacting proteins contain an RS domain, and in every case, the portion of the interacting protein that is expressed as a fusion with the activation domain of the GAL4 transcription factor contains at least part of this domain (Fig. 16B). This fact suggested that the interactions observed were mediated directly by these RS domains. In

order to verify this, constructs expressing deleted versions of USSRp58 fused to the DNA binding domain of GAL4 were prepared, co-transformed in AH109 with each of the library plasmids isolated from the screen, and tested for the activation of the reporters. Except for HCC1, all the factors showed positive interaction with the RS domain of USSRp58. Surprisingly, the C-terminus of USSRp58 seemed to mediate the interaction with HCC1. This is reminiscent of the interaction described between U2AF⁶⁵ and U2AF³⁵, where the interaction between the two RS domain-containing proteins is not mediated by their RS domains (Zhang et al., 1992).

A group of well studied splicing factors including U170K, U2AF⁶⁵, U2AF³⁵ and SF1, were tested for their ability to interact with USSRp58. The plasmids expressing these proteins as a fusion with the activation domain of GAL4 transcription factor (Davies et al., 1998) were kindly provided by Prof. Nick Hastie. As it can be seen in Fig. 17A, USSRp58 was able to interact exclusively with U2AF³⁵, and again, this interaction was mediated by its RS domain.

Next, another group of RS domain-containing proteins that have a well-characterised role in splicing was tested for their ability to interact with USSRp58. Full-length cDNAs for several members from the SR protein family were subcloned in the pACT2 vector (Clontech), and co-transformed into AH109 in combination with the different constructs expressing either full-length or deletion mutants of USSRp58 to test their ability to interact (Fig. 17B). The SR proteins SF2/ASF and SRp40 were able to interact with full-length USSRp58. These interactions seemed to be mediated by their RS domain, since they could interact with the N-terminus of USSRp58, but not with the C-terminus. In contrast, when SRp55, SRp20, 9G8 or SRp30c were expressed as preys, no interaction with USSRp58 was detected. These data indicate that USSRp58 has specificity for SF2/ASF and SRp40 RS domains, and that it does not simply bind to any RS domain. Taken together, these results

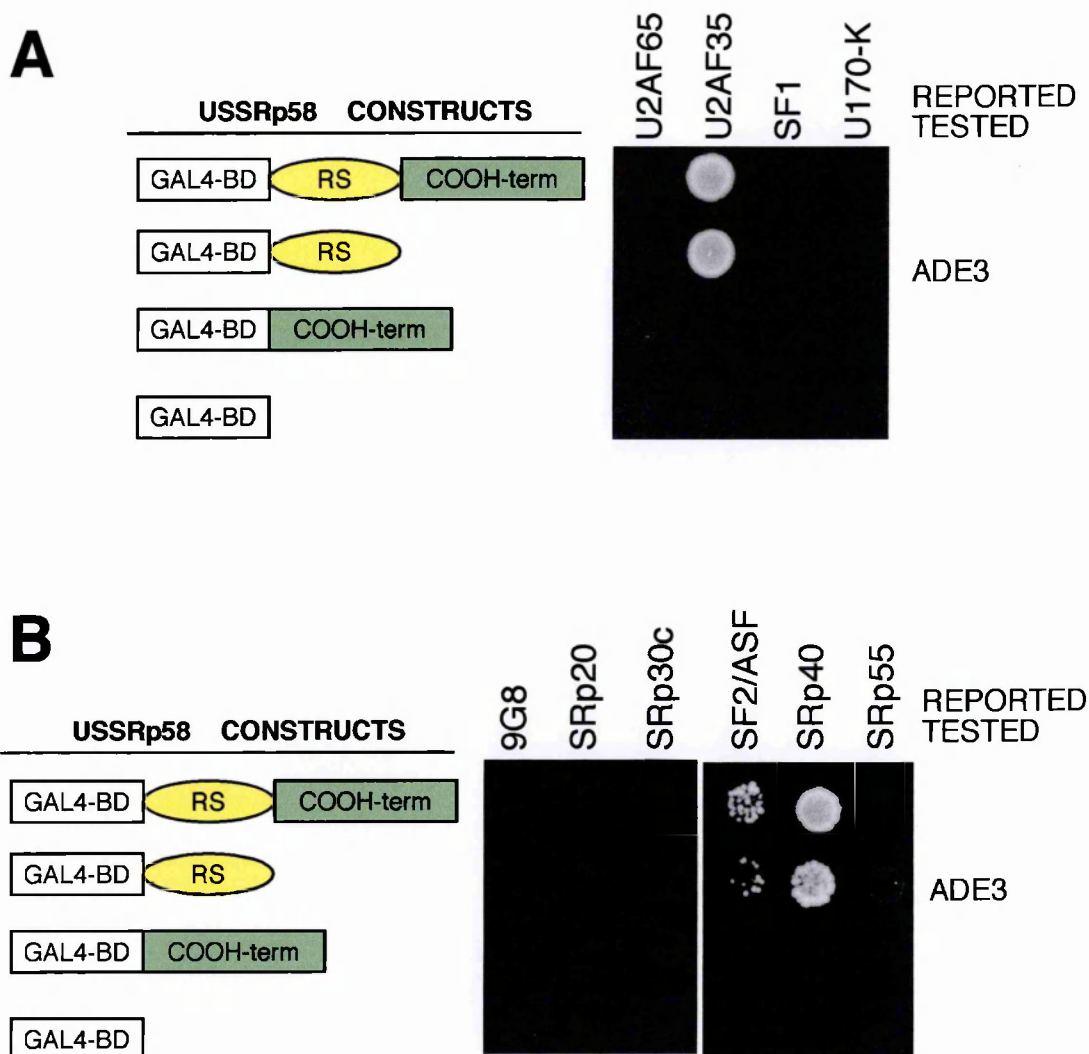


Figure 17: USSRp58 interacts with splicing factors in a yeast two-hybrid system. Interactions between wild-type or mutant versions of USSRp58 and different splicing factors were tested. **(A)** Yeast cells were co-transformed with prey plasmids encoding different splicing factors (indicated above) with bait vectors expressing either full-length USSRp58 (first row), the N-terminal half of the protein comprising the RS domain (second row), the C-terminal half of the protein (third row), or the empty vector (fourth row), and were tested for their ability to grow in the absence of Adenine. **(B)** Interactions between wild-type or mutant versions of USSRp58 and different SR proteins were tested as described in (A). In all cases shown in (A) and (B), activation of *ADE3* reporter correlated with activation of *MEL1* reporter (data not shown).

strongly suggest a role for USSRp58 in pre-mRNA splicing. Furthermore, the fact that USSRp58 interacts with U2AF³⁵ but not with U1-70K suggest a possible role for USSRp58 in 3' splice site selection.

3.1.8. USSRp58 interacts in vivo with splicing factors

I confirmed the two-hybrid interactions of USSRp58 with Luc7p and with SF2/ASF by co-immunoprecipitation (Co-IP) experiments from human 293T cell extracts. Full-length hLuc7a was subcloned in the mammalian expression vector, pCGT7 that has been previously described (Caceres et al., 1997), and following transfection into 293T cells, extracts were immunoprecipitated with a monoclonal antibody against the T7-tag. The immunoprecipitated proteins were then separated on SDS-PAGE, blotted, and detected with an anti-USSRp58 antibody. As shown in Fig. 18A, hLuc7a was able to pull down USSRp58. This interaction seemed to be specific, since USSRp58 was not pulled down when anti-T7 mAb coupled to Sheparose beads was incubated with extracts from mock-transfected cells. Furthermore, this interaction is not RNA-mediated, since treatment with ribonuclease A did not abolish this interaction.

Furthermore, SF2/ASF was detected in an immunoprecipitate from 293T cells immunoprecipitated with anti-USSRp58 antibodies (Fig. 18B). Again, this interaction was not mediated by RNA. Taken together, these results confirm the interaction data obtained with the yeast two-hybrid system and demonstrate that this novel SR-related protein functionally interacts with splicing factors.

3.1.9. USSRp58 specifically associates with a subset of snRNPs

Two-hybrid and co-immunoprecipitation data showed that USSRp58 interacts with Luc7a, a protein that is an integral component of the U1 snRNP particle in yeast, and also

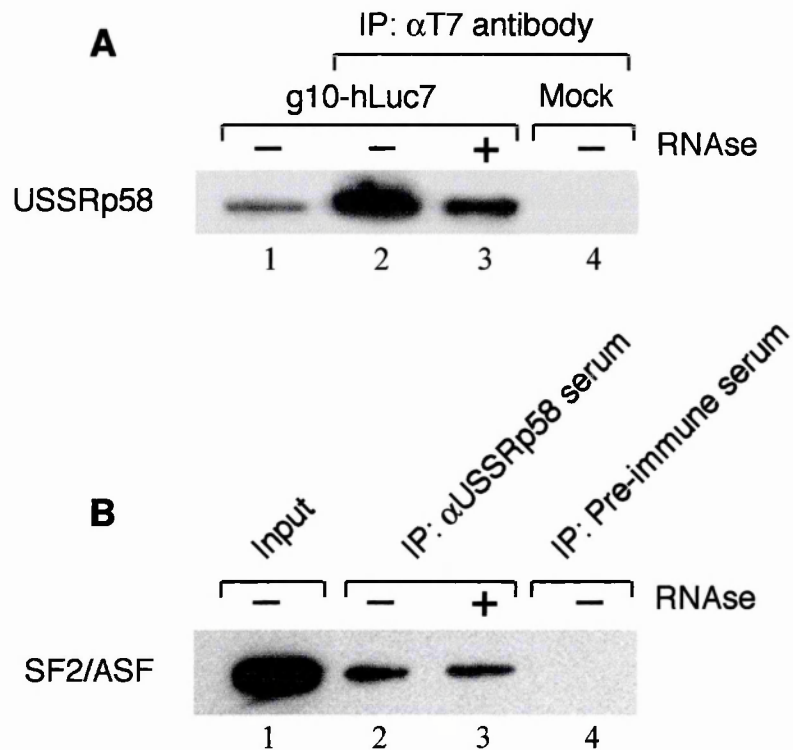


Figure 18: USSRp58 interacts with splicing factors in cultured mammalian cells.

(A) Immunoprecipitation assays with T7-hLuc7a. Extracts prepared from 293T cells either transiently transfected with a plasmid encoding T7-hLuc7a (lanes 2 and 3) or mock transfected (lane 4) were incubated with anti-T7 antibody bound to Sepharose beads (Novagen). The bound proteins were separated on a 10% SDS-polyacrylamide gel, transferred to a PVDF membrane and probed with anti-USSRp58B4 antibody. Alternatively, the immunoprecipitate was treated with RNase before loading on the gel (lane 3). Lane 1 was loaded with 2% of the amount of extract used for each immunoprecipitation. **(B)** Extracts prepared from 293T cells were incubated with either anti-USSRp58B4 antibody (lanes 2 and 3) or pre-immune serum (lane 4) bound to Sepharose beads, and analysed as described in (A). The blot was probed with mAb 96 antibody, which recognises SF2/ASF (Hanamura et al., 1998). Alternatively, the immunoprecipitate was treated with RNase before loading on the gel (lane 3). Lane 1 was loaded with 2% of the amount of extract used for each immunoprecipitation.

with HCC1, a U2AF65-like factor. Therefore, I sought to investigate whether this novel SR-related protein was associated with snRNP particles, by performing immunoprecipitations at various salt concentrations using antibodies against the USSRp58 protein. Co-precipitated RNAs were phenol-extracted, labelled with [^{32}P]pCp, fractionated by urea-PAGE and detected by autoradiography. Since U6 snRNA is not labeled efficiently with [^{32}P]pCp (Lund and Dahlberg, 1992), its presence was detected by Northern blot hybridization. For comparison, immunoprecipitations were performed with anti-USSRp58 antibody pre-absorbed with PEP2.

Surprisingly, antibodies against USSRp58 efficiently precipitated U2, U4, U5 and U6 snRNAs when the incubation and washes were done in the presence of 150 mM NaCl (Fig. 19, lane 2), suggesting that this protein is associated with these snRNPs in splicing conditions. Although some U1 snRNA is precipitated with anti-USSRp58 antibodies, it is not enriched in the precipitate compared to the precipitate obtained with Y12 antibody (Fig. 19, lane 1), as is the case for U2, U4, U5 and U6 snRNAs. When the incubation and washes were done in the presence of 250 mM NaCl, all these interactions were abolished, suggesting that the protein is not a stable component of any of these snRNPs. These interactions were specific, since no snRNAs were immunoprecipitated when the antibody was pre-absorbed with PEP2. These results demonstrate that USSRp58 is associated with a subset of snRNPs in splicing conditions. Based on this results, I decided to call the protein USSRp58, for U snRNP SR-related protein of 58 kDa. These results suggest that USSRp58 could have a role in late events in spliceosome assembly and it could also be involved in splicing catalysis as part of the active spliceosome.

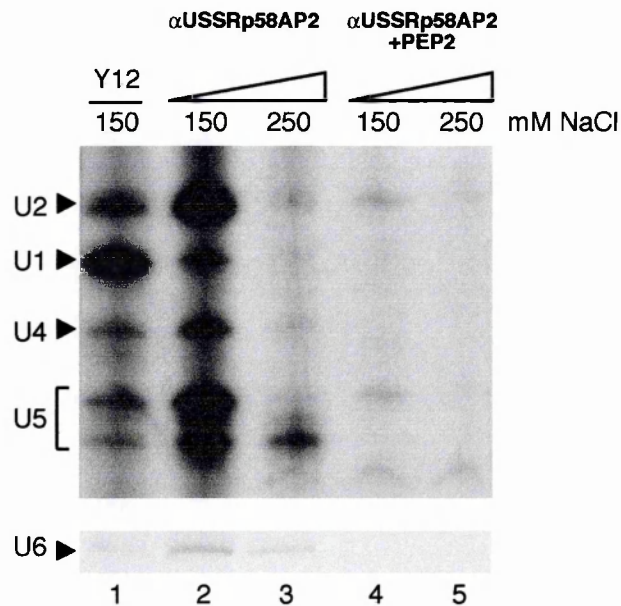


Figure 19: Immunoprecipitation of snRNPs associated with USSRp58. HeLa nuclear extracts were immunoprecipitated at the salt concentrations indicated above the lanes, with Y12 antibody (lane 1), which recognises Sm proteins, anti-USSRp58AP2 antibody (lanes 2 and 3) or anti-USSRp58AP2 antibody pre-absorbed with PEP2 (lanes 3 and 4). Co-precipitated RNAs were extracted, labeled with [^{32}P]pCp, fractionated by 10% urea-PAGE and detected by autoradiography. The positions of the snRNAs are indicated on the left. Since U6 snRNA is not labelled efficiently with [^{32}P]pCp, its presence was detected by Northern blot hybridization (lower panel).

3.1.10. USSRp58 does not complement splicing-deficient S100 HeLa extracts

SF2/ASF, the prototype member of the SR protein family was identified based on its ability to complement an otherwise inactive cytosolic HeLa S100 extract (Krainer et al., 1990b; Zahler et al., 1992). Subsequently, the ability of individual SR proteins to complement S100 extracts has been adapted as the standard assay to test their activity in constitutive splicing. In contrast to SR proteins, USSRp58 was unable to restore splicing of the β -globin splicing reporter in S100 extracts (Fig. 20A). Total SR proteins purified from HeLa cells could support the splicing of the E1A reporter when added to S100 extracts as well (Fig. 20B). When purified recombinant his-myc-USSRp58 was added to S100 extracts in the presence of bulk SR proteins, a repression in the use of the proximal 13S 5' splice site was observed, recapitulating the effect observed in the *in vivo* experiments (Figures 20B and see below, Fig. 22). These data suggest that one possible function of USSRp58 could be to negatively regulate the effect of SR proteins. However, we can not rule out at this stage, a role for this protein in later steps of splicing.

3.1.11. USSRp58 is required for pre-mRNA splicing

USSRp58 interacts with a subset of snRNPs and splicing factors like U2AF³⁵, suggesting that this protein could have a role in splicing beyond simply antagonising SR proteins. Immunodepletion of USSRp58 from nuclear extracts was performed to determine whether USSRp58 is required for splicing *in vitro*. Nuclear extracts were depleted of USSRp58 and then tested for the ability to splice a *fushi tarazu*-derived RNA substrate. A Western blot of control and immunodepleted extracts, probed with anti-USSRp58 antibody, is shown in Figure 21A. After three rounds of depletion at 1M KCl with anti-USSRp58

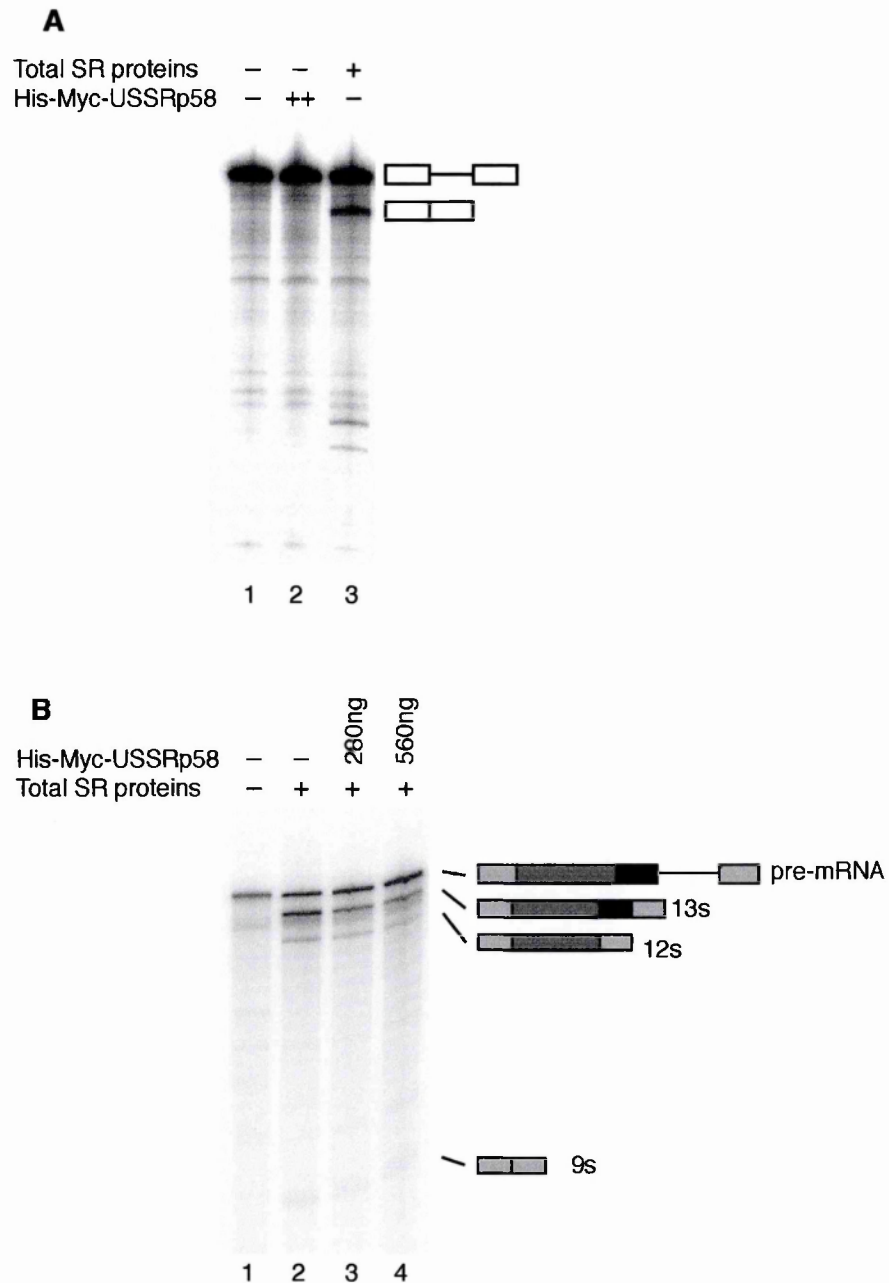


Figure 20: USSRp58 does not complement splicing-deficient S100 extracts and can regulate 5' splice site selection *in vitro*. (A) *In vitro* splicing was performed with a beta-globin slicing reporter. S100 splicing deficient extracts were supplemented with either 560 ng of His-Myc-USSRp58 (lane 2) or 250 ng of bulk SR proteins purified from HeLa cells (lane 3). The positions of pre-mRNA and spliced mRNA are indicated schematically on the right. (B) *In vitro* splicing of the E1A reporter was performed in S100 extracts supplemented with 250 ng of bulk SR proteins (lanes 2-4) and increasing amounts of His-Myc-USSRp58 (lanes 3 and 4). The positions of pre-mRNA and splice products are indicated schematically on the right.

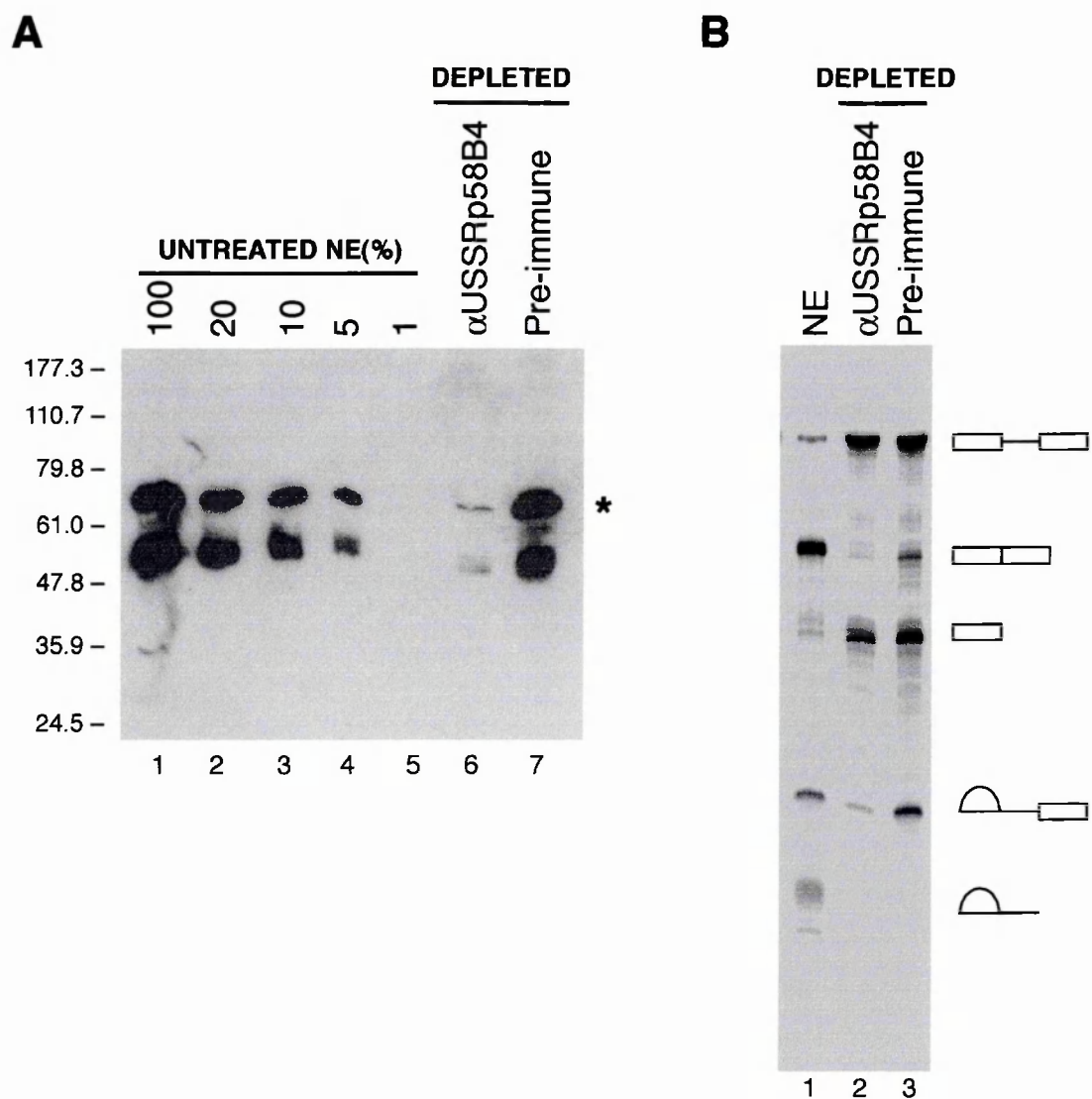


Figure 21: Effect of USSRp58 immunodepletion on pre-mRNA splicing *in vitro*. (A) The extent of USSRp58 depletion from HeLa nuclear extracts was assayed by Western blot analyses with an anti-USSRp58 antibody in serial dilutions of untreated extracts (lanes 1-5), extracts depleted with anti-USSRp58B4 antibody (lane 6) or pre-immune serum-depleted extracts (lane 7). Protein size markers (kDa) are indicated on the left. The uncharacterised protein that cross-reacts with anti-USSRp58 antibodies is indicated with an asterisk. (B) Effect of immunodepletion of USSRp58 on the splicing of *fushi tarazu* (*ftz*) splicing reporter. *Ftz* pre-mRNA was incubated in untreated nuclear extract (lane 1), USSRp58-depleted extract (lane 2) or extract depleted with pre-immune serum (lane 3). Positions of splicing intermediates and products are indicated on the right.

antibody, the concentration of the protein in the extract was reduced by at least 95%, whereas it was not significantly depleted with the pre-immune antibody. Splicing of *ftz* RNA was assayed with untreated, pre-immune depleted and anti-USSRp58 depleted extracts (Fig. 21B). Treatment with pre-immune serum resulted in a relative accumulation of intermediate products and a lower yield of spliced product when compared to the untreated extracts. These effects may, however, be related to nonspecific (e.g., dilution) effects of the depletion procedure. Nevertheless, pre-immune depleted extracts were able to carry out splicing, as it was revealed by the accumulation of spliced product. In contrast, when splicing of *ftz* RNA was assayed with USSRp58-depleted extracts, no accumulation of spliced product was observed, indicating that USSRp58 is essential for splicing. The second step of the splicing reaction, which generates released lariat intermediate and ligated exons, was preferentially affected by USSRp58 depletion under these conditions, indicating that this protein may play a role in this step. However, we could not rule out the possibility that it was required for the first step as we did not detect any accumulation of the first-step splicing intermediates.

3.1.12. USSRp58 can regulate alternative 5' site selection in vivo and in vitro

We investigated the role of USSRp58 in the regulation of alternative splicing in living cells. Modulation of alternative splice site selection can be tested *in vivo* by co-transfection of an expression vector encoding the protein of interest with an alternative spliced substrate reporter (Caceres et al., 1994). Wild-type USSRp58 was overexpressed in HeLa cells and changes in the patterns of alternative splicing of an adenovirus E1A splicing reporter, that contains three alternative 5' splice sites, were assayed (Fig. 22A). In agreement with previous results, wild-type SF2/ASF strongly activated the 13S 5' splice site (Caceres et al.,

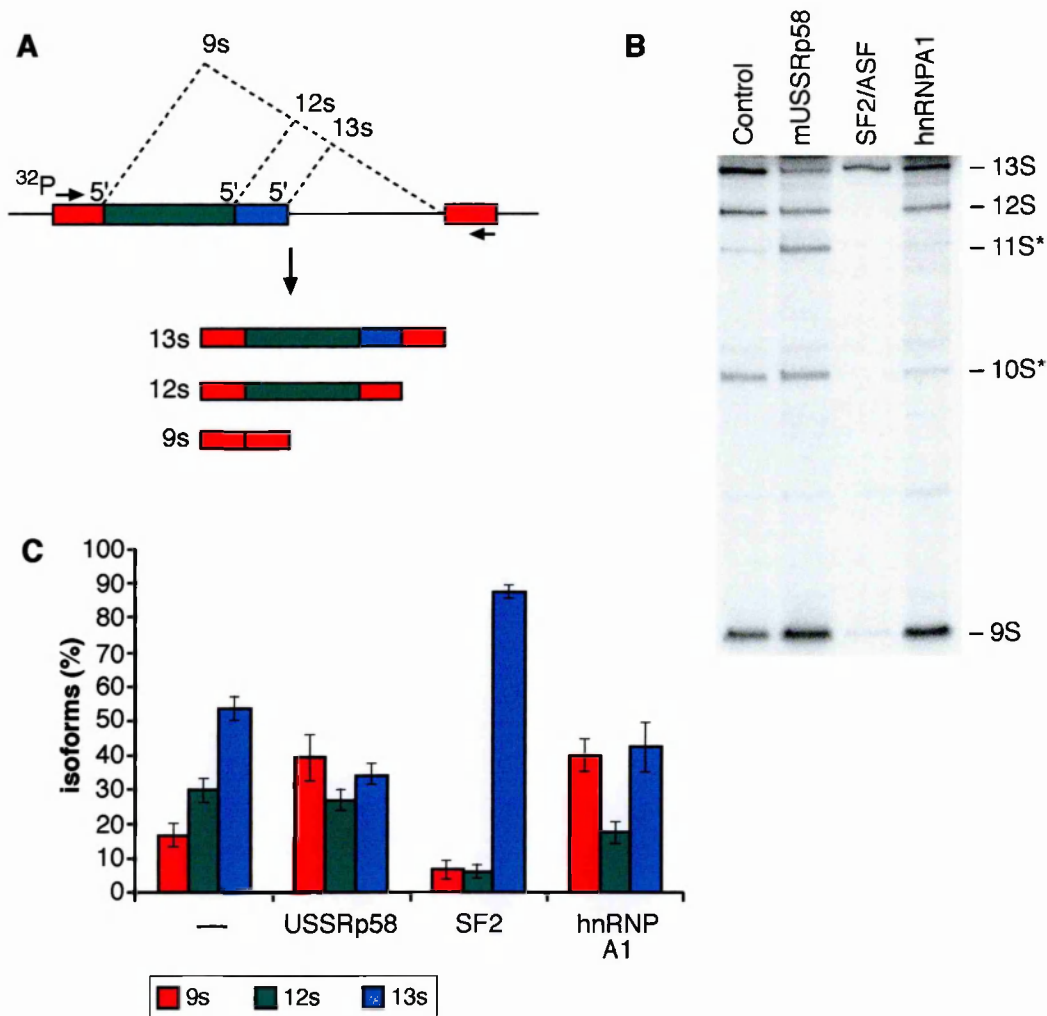


Figure 22: USSRp58 can regulate 5' splice site selection *in vivo*. *In vivo* splicing analyses were performed with HeLa cells transiently co-transfected with an adenovirus E1A reporter plasmid and expression plasmids encoding for either mouse USSRp58, hnRNP A1 or SF2/ASF. **(A)** Diagram of the E1A reporter gene. The alternative 5' splice sites and splicing events that generate 13S, 12S, and 9S mRNAs are shown schematically. The location of the exon primers used for RT-PCR analysis is shown. **(B)** HeLa cells were transiently co-transfected with the adenovirus E1A reporter plasmid and the expression constructs for each of the proteins indicated above or the parental plasmid (control). RNA was harvested at 24 hr post-transfection and analyzed by RT-PCR with a labelled forward primer, denaturing PAGE and autoradiography, as described in Materials and Methods. The positions of the unspliced pre-mRNA, and of 13S, 12S, and 9S spliced mRNAs are indicated on the right. The 10S and 11S isoforms (*) do not arise from competition between alternative 5' splice sites. **(C)** Quantitation of E1A mRNA isoforms in transfected cells. The relative amounts of 13S, 12S, and 9S E1A mRNAs were calculated from the data in (B), using a PhosphorImager, and the percentage of each isoform is shown. Each experiment was repeated four times and the data represent averages, with bars indicating standard errors.

1994). In contrast, increased expression of USSRp58 resulted in activation of the most distal 5'splice site, giving rise to the 9S isoform, as observed for hnRNP A1 (Caceres et al., 1994). Quantitation of the relative use of the E1A 5' splice sites upon overexpression of the different proteins is shown in Fig. 22C. As it was shown in section 3.1.7 and 3.1.8, USSRp58 interacts with some SR proteins, such as SF2/ASF and SRp40. Therefore, the alternative splicing modulation effect shown by USSRp58 could result from the repression of the activity of SR proteins, such as SF2/ASF, which promotes the use of the proximal alternative 5' splice site.

Taken together, these results suggest that USSRp58 is a novel SR-related protein, associated with several U snRNP particles that has a role both in constitutive and alternative splicing.

3.2. Identification of a nuclear retention signal in the RS domain of non-shuttling SR proteins

As it was mentioned in the Introduction of this thesis, a subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. The group of shuttling SR proteins include SF2/ASF, Srp20 and 9G8, whereas SC35 and SRp40 are restricted to the nucleus (Caceres et al., 1998). It was demonstrated that the shuttling proteins SRp20 and 9G8 promote the export of intronless RNAs in both mammalian cells and *Xenopus* oocytes (Huang and Steitz, 2001). More recently, these SR proteins and also SF2/ASF, which also shuttles between the nucleus and the cytoplasm, were shown to interact directly with the major receptor for bulk mRNA export to the cytoplasm, TAP/NXF1 (Huang et al., 2003), suggesting a role for SR proteins as adapters in mRNA export. SF2/ASF has also been shown to control the stability of a PKCI-1 related mRNA (Lemaire et al., 2002). At present, it is not known if shuttling SR proteins have other roles in the cytoplasmic fate of mRNAs. It has been speculated that they could have roles in translational regulation (Caceres et al., 1998). It has also been suggested that this mechanism could be used by the cell to modulate the concentration of SR proteins in the nucleus, thus possibly affecting the splicing of some mRNAs, as it has been shown for some SR proteins antagonists, like hnRNP A1 (van der Houven van Oordt et al., 2000).

The mechanism by which some SR proteins shuttle between the nucleus and the cytoplasm is not clear. By swapping the corresponding domains between shuttling and non-shuttling SR proteins, it was shown that the RS domain of shuttling SR proteins is required for shuttling (Caceres et al., 1998). For example, a chimeric protein composed of the RS

domain of SF2/ASF (a shuttling protein) and the RRM of SC35 (a non-shuttling protein), can shuttle. However, the RS domain of shuttling SR proteins is not sufficient to promote the nuclear export of these proteins because it can not confer the ability to shuttle to a nucleoplasmin core domain reporter (Caceres et al., 1998). Nevertheless, the RS domain of shuttling SR proteins can act as nuclear export signal (NLS) when fused to an RRM (Caceres et al., 1998)

The results described above suggested that some SR proteins fail to be exported from the nucleus because they either lack a NLS or because they contain a nuclear retention signal (NRS). When the RS domain of SC35 was fused to hnRNP A1 (a shuttling protein) or to SF2/ASF, the resulting fusion proteins were not able to shuttle (see Fig. 4 in Cazalla et al., 2002), suggesting that the second hypothesis is the correct one.

To determine more precisely this putative NRS in the RS domain of SC35, two C-terminal deletions were made and were tested to see if they could relieve nuclear retention. These deletion mutants lack the last 45 and 30 amino acids and were designated SC35-RS1d and SC35-RS2d, respectively (Fig. 23A). When tested in interspecies heterokaryon assay, these two mutant proteins were able to shuttle between the nucleus and the cytoplasm (Fig. 23B). Thus, deletion of the last 30 amino acids of SC35 (SC35-RS2d) relieved nuclear retention and allowed nucleo-cytoplasmic shuttling, confirming that this sequence comprises an NRS (Fig. 23C). To determine if these NRS was sufficient to confer nuclear retention to a shuttling SR protein, the last 30 amino acids of SC35 comprising the NRS were fused to SF2/ASF (Fig. 24A) and the resulting fusion protein was tested for its ability to shuttle. As it can be seen in Fig. 24B, the fusion protein does not shuttle between the nucleus and the cytoplasm, indicating that the NRS identified in the RS domain of SC35 is transferable, dominant, and sufficient to confer nuclear retention on NES-bearing proteins.

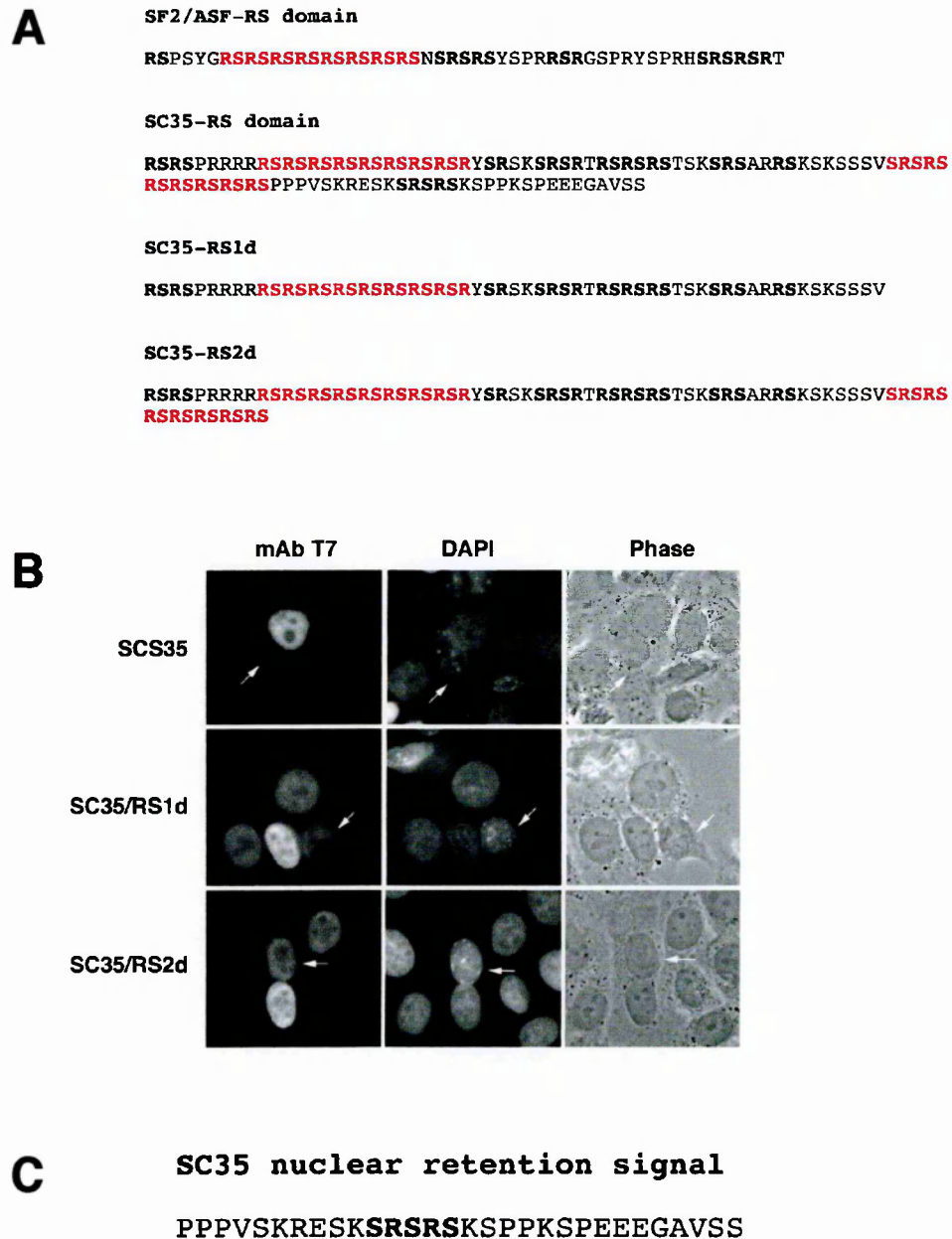


Figure 23: C-terminal deletions in the RS domain of SC35 relieve nuclear retention. (A) Sequences of the SC35 RS domain and of two C-terminal deletions. The long tracts of RS dipeptides are indicated in grey, whereas RS dipeptides outside of these tracts are indicated in bold. (B) Analysis of nucleocytoplasmic shuttling of SC35 deletion mutant proteins by transient expression in interspecies heterokaryons. Wild-type SC35 and deletion proteins were transiently expressed in HeLa cells, which were then fused with NIH 3T3 cells and analysed as described in the legend to Fig. 13. The arrows indicate the mouse nuclei within human-mouse heterokaryons. (C) Sequence of the NRS present in the RS domain of SC35.

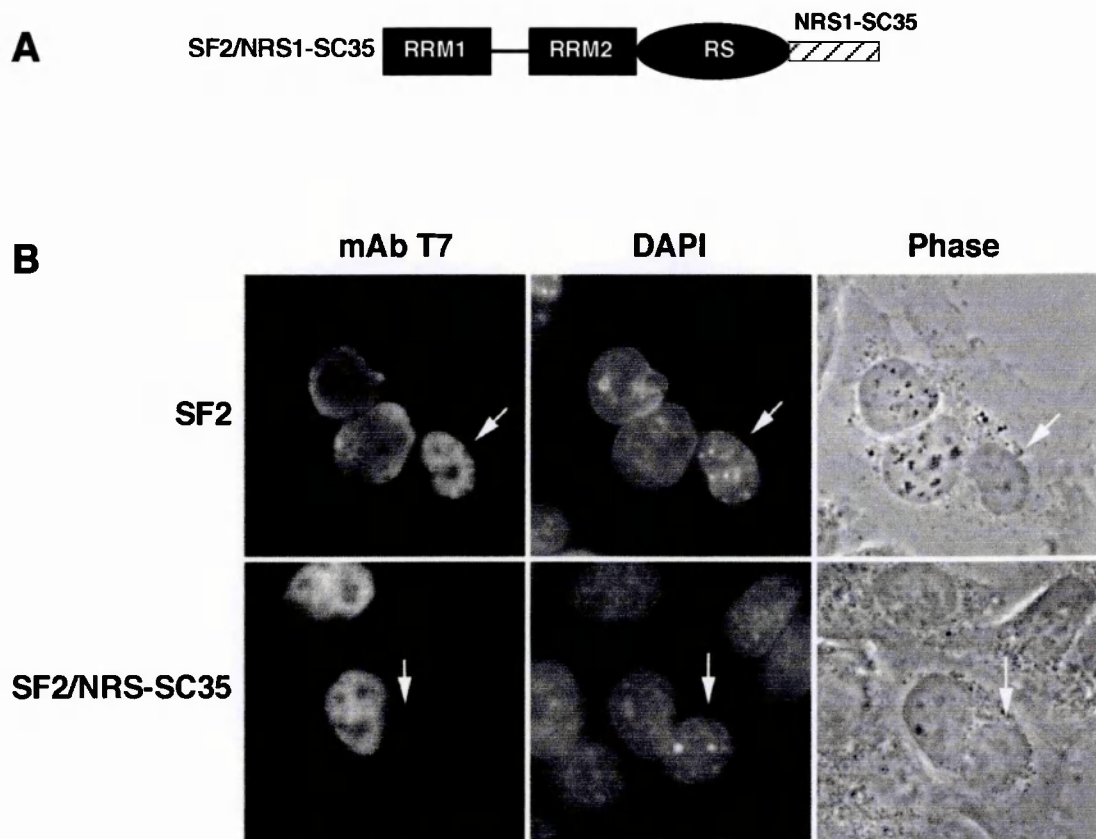


Figure 24: The NRS sequence identified in the RS domain of SC35 is transferable. (A) Diagram showing the structure of a chimeric protein in which the NRS identified in the RS domain of SC35 was fused to SF2/ASF. (B) Analysis of nucleocytoplasmic shuttling by transient expression in interspecies heterokaryons. Wild-type SF2/ASF and the chimeric protein SF2/NRS-SC35 were transiently expressed in HeLa cells, which were then fused with NIH 3T3 cells and analysed as described in the legend to Fig. 13. The arrows indicate the mouse nuclei within human-mouse heterokaryons.

Discussion

A modified gene trap strategy, based upon the subcellular localisation of the product of the trapped gene, is being used in our laboratory to identify novel genes with a potential role in RNA processing (Tate et al., 1998). Three independent avenues suggest that a novel RS-containing protein, USSRp58, has a role in pre-mRNA processing:

1. Analysis of USSRp58 sequence revealed that it contains an RS domain, which is diagnostic of factors involved in pre-mRNA splicing, and more generally in RNA processing. The subcellular localisation of the endogenous USSRp58 protein, as well of transiently expressed epitope-tagged USSRp58 protein, revealed that this protein localises to nuclear speckles, which again suggests a potential role in pre-mRNA splicing.
2. Two-hybrid analysis and immunoprecipitation/Western blot assays have been used to demonstrate that USSRp58 interacts with different proteins that have either potential or clearly defined functions in splicing. Immunoprecipitation experiments revealed that USSRp58 also interacts with spliceosomal snRNPs, indicating that it is physically associated with the splicing apparatus.
3. Using *in vivo* and *in vitro* splicing assays, USSRp58 was shown to be functionally involved in splicing. Transient overexpression of USSRp58 in cultured cells and analysis of the alternative splicing of an adenovirus E1A reporter indicated that the protein can regulate 5' splice site selection *in vivo*. Furthermore, nuclear extracts immunodepleted of USSRp58 are unable to perform splicing *in vitro*, suggesting that this protein is necessary for this process.

A final set of results describing a nuclear retention signal in the RS domain of SC35 was also included in this thesis. Chimeric proteins in which the RS domain of a non-shuttling SR protein, SC35, was fused to either hnRNP A1 or SF2/ASF (both shuttling proteins), were used to demonstrate that an NRS was present in the RS domain of SC35. Analysis of deletion mutants of SC35 indicated that the NRS is comprised in the last 30 amino acids of the protein. Finally, by fusion of the NRS found in SC35 to SF2/ASF and analysis of the shuttling properties of the fusion protein, this NRS was shown to be portable and dominant.

4.1. Structural features of USSRp58

USSRp58 is a novel protein with sequence similarity to previously identified splicing factors. It contains an RS domain, a domain found in several proteins involved in constitutive and alternative splicing regulation, such as SR proteins and SR-related proteins. RS domains have been implicated in both protein-protein and protein-RNA interactions (Amrein et al., 1994; Hedley et al., 1995; Kohtz et al., 1994; Valcarcel et al., 1996). We have shown that the RS domain of USSRp58 mediate protein-protein interactions with other RS domain-containing proteins in a yeast two-hybrid assay. The RS domain of USSRp58 showed specificity in these interactions, since it was able to interact with only two out of six members of the SR family of proteins that were tested. This protein also contains a short stretch of EK-rich sequence, which combined with the RS domain contribute extensively to the overall charge of USSRp58 and are likely responsible in part for the anomalous migration of the protein on SDS-PAGE. Another possible explanation for the difference between the predicted molecular weight of the protein and

that observed on SDS-PAGE is the presence of post-translational modifications, most likely phosphorylation of the RS domain. Accordingly, SR proteins are known to run with a higher apparent molecular mass on SDS-PAGE because of phosphorylation of the RS domain (Hanamura et al., 1998). Phosphorylation of the SR domain of USSRp58 *in vivo* remains to be demonstrated.

The unusual EK-rich region is also found in SRrp86, an SR-related protein, that was shown to regulate the activity of several SR proteins (Barnard and Patton, 2000). By analysis of chimeric proteins and deletion mutants in which the EK-rich region was removed, it was demonstrated that this domain is responsible for the inhibitory activity on splicing and splice-site selection elicited by SRrp86, most likely by titrating the activity of required splicing factors (Li et al., 2002). This could explain the effect of USSRp58 on 5' splice site selection *in vivo* and *in vitro*. Thus, the short EK-rich region present in USSRp58 could be responsible for antagonising SR proteins, affecting splice site choice in the E1A reporter. Preliminary results showed that USSRp58 is also able to antagonise SR proteins in constitutive splicing *in vitro* (data not shown). If this is the case, a mutant version of USSRp58 in which this EK-rich region is absent should not be able to enhance the choice of the distal 5' splice site in the E1A reporter and should not affect the activity of SR proteins on constitutive splicing either. The functional significance of this domain in USSRp58 will be further investigated.

USSRp58 does not contain a known RNA-binding motif. Several proteins such as Prp8, Prp16, and Tra do not contain RNA-binding motifs (Brown and Beggs, 1992; McKeown et al., 1987; Schwer and Guthrie, 1991; Zhang et al., 1992), yet they play important roles in splicing and alternative splicing regulation (Brown and Beggs, 1992; Zuo and Maniatis, 1996). Despite lacking known RNA-binding motifs, some of these splicing factors do bind to RNA. Crosslinking experiments in yeast and mammals have

demonstrated that Prp8 and its mammalian homolog p220 directly contact the pre-mRNA, lariat intermediate, and excised lariat intron (Garcia-Blanco et al., 1990; Teigelkamp et al., 1995; Whittaker and Beggs, 1991). It remains to be established whether USSRp58 is able to bind RNA, either through its RS domain or C-terminal domain.

Surprisingly, the C-terminal half of USSRp58 mediates the interaction with HCC1 in a yeast two-hybrid assay. This is reminiscent of the interaction described between the two subunits of U2AF (Zhang et al., 1992), where the RS domains present in both subunits are not involved in the interaction. Consistently, the overall domain structure of HCC1 resembles that of U2AF⁶⁵ (Imai et al., 1993). Since the function of HCC1 is not known at the moment, the significance of this interaction is unknown.

4.2. Expression and subcellular localisation of USSRp58

The expression of splicing factors is usually regulated in a tissue and developmental-specific manner. A classical example are the components in the cascade of alternative splicing events regulating the sex-determination pathway in *Drosophila melanogaster* (reviewed in Gebauer et al., 1997). Alternative splicing of a variety of pre-mRNAs can be also affected *in vivo* by the relative ratio of antagonistic splicing regulators. For example, members of the SR protein family and of the hnRNP A/B family of proteins, which antagonise each other as regulators of alternative splicing, vary naturally over a wide range in different tissues and cell lines (Hanamura et al., 1998). Northern blot analysis showed that and analysesUSSRp58 is widely yet differentially expressed across a variety of mouse adult tissues, consistent with a potential role as a splicing regulator.

Antibodies raised against peptides derived from the USSRp58 sequence cross-react with a protein of approximately 68 kDa on Western blots from nuclear extracts and mouse tissues. This protein also seems to be widely and differentially expressed across tissues. The relationship of this protein to USSRp58 is not known at the moment. This band could be either a modified form or alternatively spliced isoform of USSRp58, although there is no direct evidence to suggest that it is related to USSRp58. Western blot analysis of extracts prepared from cells transiently transfected with pCGT7-USSRp58 showed only a single band, (see Fig. 11A). Further experimentation will be required to determine the identity of this cross-reacting band.

When used in indirect immunofluorescence experiments, the antibodies raised against USSRp58 derived peptides revealed that the endogenous protein localises exclusively to the nucleus, and shows a speckled staining pattern. Transiently expressed epitope-tagged versions of USSRp58 confirmed this localisation and showed co-localisation with SC35 in splicing speckles. This is not surprising, since the RS domain of certain SR proteins has been shown to be necessary and sufficient for targeting these factors to nuclear speckles (Caceres et al., 1997; Hedley et al., 1995; Li and Bingham, 1991). The localisation of USSRp58 to nuclear speckles is highly diagnostic of its involvement of pre-mRNA splicing.

Several proteins that have been described to participate in pre-mRNA splicing in the nucleus have been shown to travel to the cytoplasm with the spliced mRNA. For some of them, cytoplasmic functions have been described. For instance, members of the hnRNP family of proteins have been reported to regulate cytoplasmic events like mRNA localisation, translation and mRNA turnover (reviewed in Shyu and Wilkinson, 2000). The splicing regulator Sxl represses the translation of *msl-2* mRNA by binding to both the 5' and 3' untranslated regions (UTR) both *in vivo* and *in vitro* (Bashaw and Baker, 1997;

Gebauer et al., 1999; Kelley et al., 1997). Members of the SR family of proteins have also been described to shuttle between the nucleus and the cytoplasm (Caceres et al., 1998), with the RS domain being necessary but not sufficient for this activity. USSRp58 shuttles between the nucleus and the cytoplasm, suggesting that this protein might play a role in post-splicing mRNA processing. Since it is not known if USSRp58 binds directly to the mRNA, an alternative possibility is that USSRp58 travels to the cytoplasm as part of protein complexes that are deposited on the mRNA. One example that supports this possibility is the SR-related protein SRm160, which does not bind directly to the pre-mRNA (Blencowe et al., 1998), but can be cross-linked near exon-exon junctions after splicing in HeLa nuclear extracts (Le Hir et al., 2000) as part of a complex called the exon-exon junction complex (EJC), that mark the position of exon-exon junctions. Preliminary experiments showed that USSRp58 co-sediments with ribosomes in sucrose gradients (data not shown), suggesting that the protein could be involved in translation. This possibility will be further investigated.

4.3. Interactions between USSRp58 and spliceosomal components

The association of USSRp58 with spliceosomal snRNPs was tested by examining the presence of snRNAs including U1, U2, U4, U5 and U6 in the immunoprecipitates from HeLa nuclear extracts formed by anti-USSRp58 antibodies. This experiment showed that U2, U4, U5, and U6, but not U1 snRNAs could be efficiently immunoprecipitated by anti-USSRp58 antibodies in the presence of 150 mM NaCl, suggesting that USSRp58 could be associated in the nucleus with these snRNPs in splicing conditions. When the immunoprecipitations were performed in the presence of 250 mM NaCl, no snRNAs were

immunoprecipitated, suggesting that USSRp58 is not stably associated with these snRNPs. Since the anti-USSRp58 antibodies can immunoprecipitate the cross-reacting 68 kDa protein with the same efficiency as that observed for USSRp58 (see Fig. 21A), the possibility that the 68 kDa protein is associated with these snRNPs can not be ruled out. Alternative methods need to be employed to determine the way these proteins associate with the different snRNPs. For example, fractionation of nuclear extracts by glycerol gradient sedimentation can be used to resolve the U1, U2, U4/U6, U5, and U4/U6-U5 snRNPs from one another. Analysis of the resulting fractions by Western blot with the anti-USSRp58 antibody will indicate how these two proteins associate with the different snRNPs. These experiments will be performed.

Systematic examination of protein-protein interactions revealed that USSRp58 can directly and specifically interact with several known splicing factors. This protein-protein interaction profile is different from that reported for other RS domain-containing proteins such as SF2/ASF, SC35 or p54. For example, SF2/ASF and SC35 can both interact with U1-70K and U2AF³⁵, but not with U2AF⁶⁵ (Wu and Maniatis, 1993), whereas p54 interacts with SF2/ASF and U2AF⁶⁵, but can not interact with U1-70K or U2AF³⁵ (Zhang and Wu, 1996). In contrast, USSRp58 interacts with SF2/ASF and U2AF³⁵, but can not interact with U1-70K or U2AF⁶⁵. This finding suggest that USSRp58 is not directly mediating interactions between components at the 5' and 3' splice sites, but rather interacts with components of the 3' splice site. As mentioned previously, USSRp58 also interacts with the U2AF⁶⁵ related factor, HCC1, in the yeast two-hybrid system. Although the function of HCC1 is unknown, it can be speculated that its function could be similar to that of U2AF⁶⁵, since it is structurally related to this factor (Imai et al., 1993). It has been speculated that mammalian cells contain multiple U2AF-like complexes (Shepard et al., 2002). If that is the case, USSRp58 could mediate the interaction between U2AF³⁵ and HCC1, if these two

proteins are unable to interact directly. It would be interesting to determine if USSRp58, HCC1 and U2AF³⁵ are part of the same complex.

USSRp58 also interacts with Luc7a, the human counterpart of Luc7p, a yeast protein shown to be a stable component of U1 snRNP (Fortes et al., 1999). Up to date, there are no reports describing the functional characterisation of mammalian Luc7a, thus, it is difficult to ascertain the functional significance of this interaction. . The fact that Luc7p is part of U1 snRNP in yeast suggests that Luc7a could be associated with U1 snRNP in mammals. USSRp58 interacts with Luc7a, yet antibodies against USSRp58 are unable to pull down U1 snRNA efficiently. This discrepancy could be explained in two different ways. One possibility is that Luc7a is not associated with U1 snRNP at the salt concentrations at which the immunoprecipitation with anti-USSRp58 antibodies was performed. The other explanation could be that USSRp58 interacts with a pool of Luc7a that is not associated with U1 snRNP. It remains to be established which explanation is correct.

4.4. The role of USSRp58 in splicing

A role for USSRp58 in splicing was first suggested by the observation that nuclear extracts depleted from USSRp58 are unable to perform splicing. The accumulation of exon 1 intermediate to similar levels as those shown by the extract depleted with the pre-immune serum would suggest that the USSRp58-depleted extracts can perform the first step of splicing. However, at present we can not rule out a role for USSRp58 during the first catalytic step. As it was mentioned before, the 68 kDa protein that cross-reacts with anti-USSRp58 antibodies is co-depleted with USSRp58 from these extracts. Therefore, the possibility that the arrest of splicing is due to the absence of the 68 kDa can not be disregarded. Addition of recombinant USSRp58 to the depleted extracts should restore full

splicing activity if USSRp58 is the only factor responsible for the splicing deficiency observed. These experiments are currently under progress. Since the depleted extracts were tested only on one pre-mRNA, it can not be concluded that the activity depleted with anti-USSRp58 antibodies is generally required for splicing. More splicing reporters need to be tested in order to establish the general requirement of this activity for splicing.

The catalytically active spliceosome for both steps of the splicing reaction is normally referred to as the C complex. This complex can be observed as a discrete band in native gels (Das and Reed, 1999). However, different configurations of the C complex must exist because the two transesterification steps require different reactive groups (see Staley and Guthrie, 1998). This is supported by the observation that in the absence of hSlu7, a protein required for the second catalytic step, splicing is arrested in between the catalytic steps in a new complex, designated the C_{ΔhSlu7} complex (Chua and Reed, 1999). One explanation for the potential requirement of USSRp58 for the second catalytic step is that the protein is essential to perform the conformational rearrangements that occur between the first and the second catalytic steps of splicing.

Genetic and biochemical data demonstrate that the second catalytic step of splicing requires a unique set of interactions that is distinct from those occurring during spliceosome assembly and the first catalytic step (reviewed in Staley and Guthrie, 1998; Umen and Guthrie, 1995). Insight into the mechanism of step II has come mainly from work done in yeast. The proteins Prp16 (Burgess et al., 1990; Couto et al., 1987; Schwer and Guthrie, 1991), Prp17 (Jones et al., 1995; Vijayraghavan et al., 1989), Prp18 (Horowitz and Abelson, 1993; Vijayraghavan and Abelson, 1990) and Slu7 (Ansari and Schwer, 1995; Frank and Guthrie, 1992; Jones et al., 1995) functions exclusively in step II, and Prp8 functions in both step I and step II (Brown and Beggs, 1992; Teigelkamp et al., 1995). Prp22, a protein involved in the release of spliced mRNA, also plays a role in step II

(Schwer and Gross, 1998). In humans, homologs of the step II factors Prp16, Prp17, Prp18 and Slu7 have been identified and shown to be general step II splicing factors (Ben Yehuda et al., 1998; Chua and Reed, 1999; Horowitz and Krainer, 1997; Lindsey and Garcia-Blanco, 1998; Zhou and Reed, 1998). Thus, splicing factors important for step II are conserved from yeast to humans. However, although splicing in yeast and mammals follows similar overall pathways, there are some notable differences. The mechanism for recognising splice sites in mammals, which allows for a wide variation in splice site sequences and for regulated alternative splicing, is more complex than that of yeast, where splice-site sequences are nearly invariant and alternative splicing is absent (Black, 1995; Horowitz and Krainer, 1994). Identification of 3' splice sites, which occurs during the second step, is therefore likely to occur by somewhat different mechanisms in mammals and yeast. Therefore, it is not surprising that mammalian specific factors are required for the second step of splicing.

After the first catalytic step has occurred, the 3' splice site needs to be identified by the spliceosome in order to proceed with the second step of splicing. In mammals, both the branch site and the polypyrimidine tract are required for both the first and second catalytic steps. (Umen and Guthrie, 1995). This requirement is easily obscured by the strong requirement for the polypyrimidine tract in spliceosome assembly prior to the first catalytic step. However, with the appropriate substrate, a role for the polypyrimidine tract in the second catalytic step has been uncovered (Reed, 1989). In yeast, the functional role in the second step of the polypyrimidine tract was also demonstrated (Patterson and Guthrie, 1991). Accordingly, a PTB associated splicing factor (PSF) has been shown to be required for the second catalytic step in mammals (Garcia-Blanco et al., 1989; Patton et al., 1993). PSF is a polypyrimidine tract binding protein associated with PTB that has only been required for the splicing of a pre-mRNA derived from the α -tropomyosin pre-mRNA. This

substrate has an unusual branch point, located far from the 3' splice site, just upstream of a second pyrimidine tract (Smith and Nadal-Ginard, 1989), thus the role of PSF as a general second step is controversial. Possibly, different factors are required to identify different 3' splice sites prior to the second catalytic step. In this scenario, a complex formed by HCC1, USSRp58 and U2AF³⁵ could have a role recognising the pyrimidine and defining the 3' splice site of certain substrates after the first catalytic splicing has occurred. Further experimentation will be required to test this hypothesis.

USSRp58 was also shown to regulate 5' splice site selection in the E1A splicing reporter *in vivo* and *in vitro*. Mechanistically, the functional antagonism shown by SF2/ASF and hnRNP A1 on 5' splice selection in this reporter (Caceres et al., 1994) is based on competitive binding to pre-mRNA. SF2/ASF interferes with hnRNP A1 binding and promotes U1 snRNP binding at all sites, resulting in the selection of the 5' proximal site. In contrast, hnRNP A1 binds co-operatively and indiscriminately to the E1A reporter repressing the U1 snRNP binding at all sites, resulting in a shift to the distal 5' splice site (Eperon et al., 1993; Eperon et al., 2000). USSRp58 promotes the selection of the distal 5' splice sites both *in vivo* and *in vitro*, resembling the activity shown by hnRNP A1. Since USSRp58 does not interact directly with U1 snRNP as suggested by the protein-protein and protein-RNA interaction profiles described in this thesis, a direct role in the regulation of U1 snRNP binding to the different 5' splice sites seems unlikely. USSRp58 could instead promote selection of the distal 5' splice site in the E1A reporter by titrating SR proteins that were shown to interact with USSRp58, such as SF2/ASF. It is also possible that the short EK-rich present in USSRp58 might play a prominent role in the repression of the activity of such SR proteins. Further experimentation will be needed to validate this hypothesis.

4.5. Nuclear retention signals in the RS domains of non-shuttling SR proteins

Results presented in this thesis showed that SC35 is restricted to the nucleus not because it lacks a nuclear export signal (NES) but because it contains a nuclear retention signal (NRS). There is little information about nuclear retention of proteins. Specific sequences that mediate the retention of proteins in the nucleus have been described, such as the one that directs the localisation of the shuttling protein nucleolin to the nucleolus (Schmidt-Zachmann et al., 1993) or a region of approximately 200 amino acids that mediates the nuclear retention of the La protein (Simons et al., 1996). However, in none of these cases has it been shown whether these proteins have transferable NRSs capable of conferring nuclear retention on proteins that contain NESs.

An NRS was identified in hnRNP C1, a non-shuttling hnRNP (Nakielnny and Dreyfuss, 1996), supporting the idea that non-shuttling hnRNPs are confined to the nucleus due to signals that actively promote their nuclear retention. The hnRNP C1 NRS was the first sequence shown to retain NES-bearing proteins in the nucleus. It comprises a core NRS sequence of 58 amino acids that requires approximately 10 amino acids at both the amino and the carboxyl termini for full NRS function. This proline-rich NRS sequence contains clusters of basic residues, potential phosphorylation sites for casein kinase II and protein kinase C, and a potential glycosylation site (Nakielnny and Dreyfuss, 1996).

Carboxyl terminal deletions in the RS domain of SC35 showed that the NRS is contained within a 30-amino acid segment at the C-terminus of the RS domain, comprising a proline-rich sequence. The human SC35 NRS sequence is portable and dominant, since it confers nuclear retention to the shuttling protein SF2/ASF. Other than this proline-rich

composition, there is no discernible sequence similarity between the NRS sequences identified in hnRNP C1 and SC35. Both sequences are proline and serine rich and basic in character, at least in the absence of phosphorylation. SRp40, which is also retained in the nucleus and has multiple consecutive RS repeats (Caceres et al., 1998), does not contain a sequence with obvious similarity to the SC35 NRS. Interestingly, the *Drosophila* homolog of human SC35, which unlike its human counterpart has a short RS domain and shuttles between the nucleus and the cytoplasm, lacks this NRS sequence (Jamal Tazi, personal communication). Most likely, nuclear retention is mediated by interactions of the NRS with nuclear components which are yet to be defined.

4.6. Future perspectives

Many questions remain open about the possible roles of USSRp58 in pre-mRNA splicing. Further analysis of the roles of the RS and the EK-rich domains in mediating interactions with SR proteins are required. It also remains to be established whether the RS domain of USSRp58 is phosphorylated *in vivo*, and how this phosphorylation affects its activity.

It would also very interesting to investigate the significance of the shuttling activity of USSRp58 and the possible roles of this protein in the cytoplasm. Fractionation of cytoplasmic extracts by sucrose gradient sedimentation and analysis of the fractions by Western blot with the anti-USSRp58 antibody have shown that USSRp58 co-sediments with 80S ribosomes and with polysomes, suggesting a role involved in translation of mRNAs in the cytoplasm (data not shown). Even though the yeast two-hybrid screen did not reveal interactions with known translation factors, USSRp58 interacts with SR proteins,

which have a role in translation (Jeremy Sanford and Javier Caceres, unpublished results). The functional significance of the association of USSRp58 with the translation machinery will be pursued.

It would also be very interesting to determine gene targets which expression can be affected by USSRp58. A good approach that could be used to identify mRNAs associated with USSRp58 is to perform immunoprecipitation experiments with anti-USSRp58 antibodies from nuclear and cytoplasmic fractions followed by identification of the mRNAs present in the immunoprecipitates by microarray analyses.

One essential question that needs to be answered is the identity of the 68 kDa protein that cross-reacts with anti-USSRp58 antibodies. Since this protein is efficiently immunoprecipitated by anti-USSRp58 antibodies, immunoprecipitation experiments coupled with SDS-PAGE fractionation and mass spec analysis are currently being performed to assess the identity of this protein. Another question that remains unexplored is the ability of USSRp58 to bind directly to RNA. Cross-linking experiments *in vivo* and *in vitro* will be performed to elucidate this possibility.

The potential requirement and function of USSRp58 during the second catalytic step of splicing is perhaps the most intriguing question remaining open. Is USSRp58 necessary for the spliceosome to undergo the conformational changes occurring during the second catalytic step? Does USSRp58 play a role in the selection of the 3' splice site prior to the step II of splicing? If this is the case, what is the significance of its interaction with HCC1 and U2AF³⁵? Are there complexes analogous to U2AF needed for the recognition of the 3' splice site during the second catalytic step of splicing? Great effort will be put in answering these questions in the near future.

Finally, it would be interesting to determine the nuclear factors that could mediate the retention of SC35 in the nucleus. A yeast two-hybrid screen using the NRS found in SC35

as a bait is currently being performed to answer this question. It was shown that SC35 can be cross-linked to poly(A)⁺ RNA in the nucleus but not in the cytoplasm, indicating that SC35 is removed from the spliced mRNA before it exits the nucleus (Huang and Steitz, 2001). It would be also important to define whether the NRS plays a role in the selection of SC35 as a protein to be removed from spliced mRNAs before they leave the nucleus. It has also been shown that shuttling SR proteins interact with TAP/NXF1 (Huang et al., 2003), functioning as adapters proteins for TAP-dependent mRNA export. It would be important to determine whether the NRS found in the RS domain of SC35 has a role in preventing the interaction of SC35 with TAP/NFX1.

References

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185-2195.
- Amrein, H., Hedley, M. L., and Maniatis, T. (1994). The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell* 76, 735-746.
- Ansari, A., and Schwer, B. (1995). SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *Embo J* 14, 4001-4009.
- Barnard, D. C., and Patton, J. G. (2000). Identification and characterization of a novel serine-arginine-rich splicing regulatory protein. *Mol Cell Biol* 20, 3049-3057.
- Bashaw, G. J., and Baker, B. S. (1997). The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* 89, 789-798.
- Ben Yehuda, S., Dix, I., Russell, C. S., Levy, S., Beggs, J. D., and Kupiec, M. (1998). Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. *Rna* 4, 1304-1312.
- Bendixen, C., Gangloff, S., and Rothstein, R. (1994). A yeast mating-selection scheme for detection of protein-protein interactions. *Nucleic Acids Res* 22, 1778-1779.
- Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol Cell Biol* 12, 3165-3175.
- Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol* 11, 347-351.
- Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol* 14, 336-342.
- Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem* 270, 2411-2414.
- Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74, 3171-3175.
- Berglund, J. A., Fleming, M. L., and Rosbash, M. (1998). The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *Rna* 4, 998-1006.

- Birney, E., Kumar, S., and Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 21, 5803-5816.
- Black, D. L. (1995). Finding splice sites within a wilderness of RNA. *Rna* 1, 763-771.
- Blencowe, B. J., Issner, R., Nickerson, J. A., and Sharp, P. A. (1998). A coactivator of pre-mRNA splicing. *Genes Dev* 12, 996-1009.
- Boucher, L., Ouzounis, C. A., Enright, A. J., and Blencowe, B. J. (2001). A genome-wide survey of RS domain proteins. *Rna* 7, 1693-1701.
- Brosi, R., Groning, K., Behrens, S. E., Luhrmann, R., and Kramer, A. (1993a). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* 262, 102-105.
- Brosi, R., Hauri, H. P., and Kramer, A. (1993b). Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem* 268, 17640-17646.
- Brown, J. D., and Beggs, J. D. (1992). Roles of PRP8 protein in the assembly of splicing complexes. *Embo J* 11, 3721-3729.
- Bruzik, J. P., and Maniatis, T. (1995). Enhancer-dependent interaction between 5' and 3' splice sites in trans. *Proc Natl Acad Sci U S A* 92, 7056-7059.
- Burge, C., Tuschl, T., and Sharp, P. (1999). Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World*, R. F. Gesteland, T. R. Chech, and H. F. Atkins, eds. (New York, Cold Spring Harbor Laboratory Press), pp. 525-560.
- Burgess, S., Couto, J. R., and Guthrie, C. (1990). A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell* 60, 705-717.
- Caceres, J. F., and Kornblihtt, A. R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet* 18, 186-193.
- Caceres, J. F., and Krainer, A. R. (1993). Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *Embo J* 12, 4715-4726.
- Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., and Krainer, A. R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* 138, 225-238.
- Caceres, J. F., Screaton, G. R., and Krainer, A. R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev* 12, 55-66.
- Caceres, J. F., Stamm, S., Helfman, D. M., and Krainer, A. R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* 265, 1706-1709.

- Cao, W., and Garcia-Blanco, M. A. (1998). A serine/arginine-rich domain in the human U1 70k protein is necessary and sufficient for ASF/SF2 binding. *J Biol Chem* 273, 20629-20635.
- Cao, W., Jamison, S. F., and Garcia-Blanco, M. A. (1997). Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *Rna* 3, 1456-1467.
- Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *Rna* 5, 468-483.
- Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R., and Stevenin, J. (1994). Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *Embo J* 13, 2639-2649.
- Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A. R., and Cáceres, J. F. (2002). Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol* 22, 6871-6882.
- Cecconi, F., and Meyer, B. I. (2000). Gene trap: a way to identify novel genes and unravel their biological function. *FEBS Lett* 480, 63-71.
- Chaudhary, N., McMahon, C., and Blobel, G. (1991). Primary structure of a human arginine-rich nuclear protein that colocalizes with spliceosome components. *Proc Natl Acad Sci U S A* 88, 8189-8193.
- Chiara, M. D., and Reed, R. (1995). A two-step mechanism for 5' and 3' splice-site pairing. *Nature* 375, 510-513.
- Chien, C. T., Bartel, P. L., Sternglanz, R., and Fields, S. (1991). The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* 88, 9578-9582.
- Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8.
- Chua, K., and Reed, R. (1999). Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. *Genes Dev* 13, 841-850.
- Colot, H. V., Stutz, F., and Rosbash, M. (1996). The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev* 10, 1699-1708.
- Colwill, K., Feng, L. L., Yeakley, J. M., Gish, G. D., Cáceres, J. F., Pawson, T., and Fu, X. D. (1996a). SRPK1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. *J Biol Chem* 271, 24569-24575.

- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996b). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *Embo J* 15, 265-275.
- Couto, J. R., Tamm, J., Parker, R., and Guthrie, C. (1987). A trans-acting suppressor restores splicing of a yeast intron with a branch point mutation. *Genes Dev* 1, 445-455.
- Cowper, A. E., Caceres, J. F., Mayeda, A., and Sreaton, G. R. (2001). Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing. *J Biol Chem* 276, 48908-48914.
- Cramer, P., Caceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E., and Kornblihtt, A. R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell* 4, 251-258.
- Crispino, J. D., Blencowe, B. J., and Sharp, P. A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* 265, 1866-1869.
- Crispino, J. D., and Sharp, P. A. (1995). A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes Dev* 9, 2314-2323.
- Das, R., and Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *Rna* 5, 1504-1508.
- Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* 5, 779-787.
- Dauksaite, V., and Akusjarvi, G. (2002). Human splicing factor ASF/SF2 encodes for a repressor domain required for its inhibitory activity on pre-mRNA splicing. *J Biol Chem* 277, 12579-12586.
- Dauwalder, B., and Mattox, W. (1998). Analysis of the functional specificity of RS domains in vivo. *Embo J* 17, 6049-6060.
- Davies, R. C., Calvio, C., Bratt, E., Larsson, S. H., Lamond, A. I., and Hastie, N. D. (1998). WT1 interacts with the splicing factor U2AF65 in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev* 12, 3217-3225.
- Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 3, 195-205.
- Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62, 289-321.
- Du, C., McGuffin, M. E., Dauwalder, B., Rabinow, L., and Mattox, W. (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. *Mol Cell* 2, 741-750.

- Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* *419*, 86-90.
- Duncan, P. I., Stojdl, D. F., Marius, R. M., and Bell, J. C. (1997). In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Mol Cell Biol* *17*, 5996-6001.
- Duncan, P. I., Stojdl, D. F., Marius, R. M., Scheit, K. H., and Bell, J. C. (1998). The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing. *Exp Cell Res* *241*, 300-308.
- Eldridge, A. G., Li, Y., Sharp, P. A., and Blencowe, B. J. (1999). The SRm160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci U S A* *96*, 6125-6130.
- Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A., and Krainer, A. R. (1993). Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *Embo J* *12*, 3607-3617.
- Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G., and Krainer, A. R. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* *20*, 8303-8318.
- Estmer Nilsson, C., Petersen-Mahrt, S., Durot, C., Shtrichman, R., Krainer, A. R., Kleinberger, T., and Akusjarvi, G. (2001). The adenovirus E4-ORF4 splicing enhancer protein interacts with a subset of phosphorylated SR proteins. *Embo J* *20*, 864-871.
- Etzerodt, M., Vignali, R., Ciliberto, G., Scherly, D., Mattaj, I. W., and Philipson, L. (1988). Structure and expression of a *Xenopus* gene encoding an snRNP protein (U1 70K). *Embo J* *7*, 4311-4321.
- Evans, M. J., Carlton, M. B., and Russ, A. P. (1997). Gene trapping and functional genomics. *Trends Genet* *13*, 370-374.
- Fetzer, S., Lauber, J., Will, C. L., and Luhrmann, R. (1997). The [U4/U6.U5] tri-snRNP-specific 27K protein is a novel SR protein that can be phosphorylated by the snRNP-associated protein kinase. *Rna* *3*, 344-355.
- Fleckner, J., Zhang, M., Valcarcel, J., and Green, M. R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* *11*, 1864-1872.
- Fortes, P., Bilbao-Cortes, D., Fornerod, M., Rigaut, G., Raymond, W., Seraphin, B., and Mattaj, I. W. (1999). Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. *Genes Dev* *13*, 2425-2438.
- Frank, D., and Guthrie, C. (1992). An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev* *6*, 2112-2124.

- Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. *Rna* 1, 663-680.
- Fu, X. D., and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* 343, 437-441.
- Fu, X. D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proc Natl Acad Sci U S A* 89, 1725-1729.
- Gama-Carvalho, M., and Carmo-Fonseca, M. (2001). The rules and roles of nucleocytoplasmic shuttling proteins. *FEBS Lett* 498, 157-163.
- Garcia-Blanco, M. A., Anderson, G. J., Beggs, J., and Sharp, P. A. (1990). A mammalian protein of 220 kDa binds pre-mRNAs in the spliceosome: a potential homologue of the yeast PRP8 protein. *Proc Natl Acad Sci U S A* 87, 3082-3086.
- Garcia-Blanco, M. A., Jamison, S. F., and Sharp, P. A. (1989). Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev* 3, 1874-1886.
- Ge, H., and Manley, J. L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* 62, 25-34.
- Ge, H., Zuo, P., and Manley, J. L. (1991). Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* 66, 373-382.
- Gebauer, F., Corona, D. F., Preiss, T., Becker, P. B., and Hentze, M. W. (1999). Translational control of dosage compensation in *Drosophila* by Sex-lethal: cooperative silencing via the 5' and 3' UTRs of msl-2 mRNA is independent of the poly(A) tail. *Embo J* 18, 6146-6154.
- Gebauer, F., Merendino, L., Hentze, M. W., and Valcarcel, J. (1997). Novel functions for 'nuclear factors' in the cytoplasm: the Sex-lethal paradigm. *Semin Cell Dev Biol* 8, 561-566.
- Gottschalk, A., Tang, J., Puig, O., Salgado, J., Neubauer, G., Colot, H. V., Mann, M., Seraphin, B., Rosbash, M., Luhrmann, R., and Fabrizio, P. (1998). A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *Rna* 4, 374-393.
- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 18, 4752-4760.
- Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *Rna* 6, 1197-1211.
- Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 17, 100-107.

- Graveley, B. R., Hertel, K. J., and Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *Rna* 7, 806-818.
- Graveley, B. R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol Cell* 1, 765-771.
- Gross, T., Richert, K., Mierke, C., Lutzberger, M., and Kaufer, N. F. (1998). Identification and characterization of *srp1*, a gene of fission yeast encoding a RNA binding domain and a RS domain typical of SR splicing factors. *Nucleic Acids Res* 26, 505-511.
- Gui, J. F., Tronchere, H., Chandler, S. D., and Fu, X. D. (1994). Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc Natl Acad Sci U S A* 91, 10824-10828.
- Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr., and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *Rna* 4, 430-444.
- Hannon, G. J., Maroney, P. A., and Nilsen, T. W. (1991). U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro. *J Biol Chem* 266, 22792-22795.
- Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A* 99, 16719-16724.
- Hastings, M. L., and Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* 13, 302-309.
- Hedley, M. L., Amrein, H., and Maniatis, T. (1995). An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc Natl Acad Sci U S A* 92, 11524-11528.
- Hoffman, B. E., and Lis, J. T. (2000). Pre-mRNA splicing by the essential *Drosophila* protein B52: tissue and target specificity. *Mol Cell Biol* 20, 181-186.
- Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., and Keene, J. D. (1991). RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci U S A* 88, 2495-2499.
- Horowitz, D. S., and Abelson, J. (1993). A U5 small nuclear ribonucleoprotein particle protein involved only in the second step of pre-mRNA splicing in *Saccharomyces cerevisiae*. *Mol Cell Biol* 13, 2959-2970.
- Horowitz, D. S., and Krainer, A. R. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends Genet* 10, 100-106.

- Horowitz, D. S., and Krainer, A. R. (1997). A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes Dev* 11, 139-151.
- Huang, Y., Gattoni, R., Stevenin, J., and Steitz, J. A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell* 11, 837-843.
- Huang, Y., and Steitz, J. A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell* 7, 899-905.
- Imai, H., Chan, E. K., Kiyosawa, K., Fu, X. D., and Tan, E. M. (1993). Novel nuclear autoantigen with splicing factor motifs identified with antibody from hepatocellular carcinoma. *J Clin Invest* 92, 2419-2426.
- Jones, M. H., Frank, D. N., and Guthrie, C. (1995). Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. *Proc Natl Acad Sci U S A* 92, 9687-9691.
- Jumaa, H., Wei, G., and Nielsen, P. J. (1999). Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. *Curr Biol* 9, 899-902.
- Jurica, M. S., Licklider, L. J., Gygi, S. R., Grigorieff, N., and Moore, M. J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *Rna* 8, 426-439.
- Jurica, M. S., and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12, 5-14.
- Kadener, S., Cramer, P., Nogues, G., Cazalla, D., de la Mata, M., Fededa, J. P., Werbach, S. E., Srebrow, A., and Kornblihtt, A. R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *Embo J* 20, 5759-5768.
- Kadener, S., Fededa, J. P., Rosbash, M., and Kornblihtt, A. R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc Natl Acad Sci U S A* 99, 8185-8190.
- Kandels-Lewis, S., and Seraphin, B. (1993). Involvement of U6 snRNA in 5' splice site selection. *Science* 262, 2035-2039.
- Kanopka, A., Muhlemann, O., and Akusjarvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381, 535-538.
- Kanopka, A., Muhlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C., and Akusjarvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* 393, 185-187.
- Kataoka, N., Bachorik, J. L., and Dreyfuss, G. (1999). Transportin-SR, a nuclear import receptor for SR proteins. *J Cell Biol* 145, 1145-1152.

- Katsu, R., Onogi, H., Wada, K., Kawaguchi, Y., and Hagiwara, M. (2002). Novel SR-rich-related protein clasp specifically interacts with inactivated Clk4 and induces the exon EB inclusion of Clk. *J Biol Chem* 277, 44220-44228.
- Kelley, R. L., Wang, J., Bell, L., and Kuroda, M. I. (1997). Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* 387, 195-199.
- Kenan, D. J., Query, C. C., and Keene, J. D. (1991). RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci* 16, 214-220.
- Kim, Y. J., and Baker, B. S. (1993). Isolation of RRM-type RNA-binding protein genes and the analysis of their relatedness by using a numerical approach. *Mol Cell Biol* 13, 174-183.
- Kim, Y. J., Zuo, P., Manley, J. L., and Baker, B. S. (1992). The *Drosophila* RNA-binding protein RBP1 is localized to transcriptionally active sites of chromosomes and shows a functional similarity to human splicing factor ASF/SF2. *Genes Dev* 6, 2569-2579.
- Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., and Manley, J. L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368, 119-124.
- Krainer, A. R., Conway, G. C., and Kozak, D. (1990a). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* 62, 35-42.
- Krainer, A. R., Conway, G. C., and Kozak, D. (1990b). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev* 4, 1158-1171.
- Krainer, A. R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* 42, 725-736.
- Krainer, A. R., Maniatis, T., Ruskin, B., and Green, M. R. (1984). Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* 36, 993-1005.
- Krainer, A. R., Mayeda, A., Kozak, D., and Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell* 66, 383-394.
- Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* 65, 367-409.
- Krauss, S. W., Larabell, C. A., Lockett, S., Gascard, P., Penman, S., Mohandas, N., and Chasis, J. A. (1997). Structural protein 4.1 in the nucleus of human cells: dynamic rearrangements during cell division. *J Cell Biol* 137, 275-289.
- Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* 11, 363-371.

- Kuroyanagi, N., Onogi, H., Wakabayashi, T., and Hagiwara, M. (1998). Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles. *Biochem Biophys Res Commun* 242, 357-364.
- Labourier, E., Bourbon, H. M., Gallouzi, I. E., Fostier, M., Allemand, E., and Tazi, J. (1999). Antagonism between RSF1 and SR proteins for both splice-site recognition in vitro and *Drosophila* development. *Genes Dev* 13, 740-753.
- Lai, M. C., Lin, R. I., Huang, S. Y., Tsai, C. W., and Tarn, W. Y. (2000). A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J Biol Chem* 275, 7950-7957.
- Lai, M. C., Lin, R. I., and Tarn, W. Y. (2001). Transportin-SR2 mediates nuclear import of phosphorylated SR proteins. *Proc Natl Acad Sci U S A* 98, 10154-10159.
- Lai, M. C., Teh, B. H., and Tarn, W. Y. (1999). A human papillomavirus E2 transcriptional activator. The interactions with cellular splicing factors and potential function in pre-mRNA processing. *J Biol Chem* 274, 11832-11841.
- Lallena, M. J., Chalmers, K. J., Llamazares, S., Lamond, A. I., and Valcarcel, J. (2002). Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109, 285-296.
- Lallena, M. J., Martinez, C., Valcarcel, J., and Correas, I. (1998). Functional association of nuclear protein 4.1 with pre-mRNA splicing factors. *J Cell Sci* 111 (Pt 14), 1963-1971.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Le Hir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *Embo J* 19, 6860-6869.
- Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J. L., and Lafyatis, R. (2002). Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev* 16, 594-607.
- Lerner, M. R., and Steitz, J. A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A* 76, 5495-5499.
- Lewis, J. D., Gunderson, S. I., and Mattaj, I. W. (1995). The influence of 5' and 3' end structures on pre-mRNA metabolism. *J Cell Sci Suppl* 19, 13-19.
- Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* 10, 1683-1698.

- Li, H., and Bingham, P. M. (1991). Arginine/serine-rich domains of the su(wa) and tra RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* 67, 335-342.
- Li, J., Barnard, D. C., and Patton, J. G. (2002). A unique glutamic acid-lysine (EK) domain acts as a splicing inhibitor. *J Biol Chem* 277, 39485-39492.
- Lindsey, L. A., and Garcia-Blanco, M. A. (1998). Functional conservation of the human homolog of the yeast pre-mRNA splicing factor Prp17p. *J Biol Chem* 273, 32771-32775.
- Liu, H. X., Zhang, M., and Krainer, A. R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev* 12, 1998-2012.
- Longman, D., Johnstone, I. L., and Caceres, J. F. (2000). Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *Embo J* 19, 1625-1637.
- Longman, D., McGarvey, T., McCracken, S., Johnstone, I. L., Blencowe, B. J., and Caceres, J. F. (2001). Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of *C. elegans*. *Curr Biol* 11, 1923-1933.
- Lund, E., and Dahlberg, J. E. (1992). Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. *Science* 255, 327-330.
- Luo, M. J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A* 96, 14937-14942.
- Lutzberger, M., Gross, T., and Kaufer, N. F. (1999). Srp2, an SR protein family member of fission yeast: in vivo characterization of its modular domains. *Nucleic Acids Res* 27, 2618-2626.
- Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293, 1836-1839.
- Lynch, K. W., and Maniatis, T. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes Dev* 10, 2089-2101.
- MacMillan, A. M., Query, C. C., Allerson, C. R., Chen, S., Verdine, G. L., and Sharp, P. A. (1994). Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev* 8, 3008-3020.
- Madhani, H. D., and Guthrie, C. (1994). Randomization-selection analysis of snRNAs in vivo: evidence for a tertiary interaction in the spliceosome. *Genes Dev* 8, 1071-1086.
- Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205-2208.

- Makarova, O. V., Makarov, E. M., and Luhrmann, R. (2001). The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *Embo J* 20, 2553-2563.
- Mancebo, R., Lo, P. C., and Mount, S. M. (1990). Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. *Mol Cell Biol* 10, 2492-2502.
- Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.
- Manley, J. L., and Tacke, R. (1996). SR proteins and splicing control. *Genes Dev* 10, 1569-1579.
- Maroney, P. A., Hannon, G. J., and Nilsen, T. W. (1990). Transcription and cap trimethylation of a nematode spliced leader RNA in a cell-free system. *Proc Natl Acad Sci U S A* 87, 709-713.
- Maroney, P. A., Yu, Y. T., Jankowska, M., and Nilsen, T. W. (1996). Direct analysis of nematode cis- and trans-spliceosomes: a functional role for U5 snRNA in spliced leader addition trans-splicing and the identification of novel Sm snRNPs. *Rna* 2, 735-745.
- Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M. Q., Gardiner, E. M., and Krainer, A. R. (1999a). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *Embo J* 18, 4560-4570.
- Mayeda, A., Helfman, D. M., and Krainer, A. R. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol Cell Biol* 13, 2993-3001.
- Mayeda, A., and Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68, 365-375.
- Mayeda, A., Screaton, G. R., Chandler, S. D., Fu, X. D., and Krainer, A. R. (1999b). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol Cell Biol* 19, 1853-1863.
- McCracken, S., Lambermon, M., and Blencowe, B. J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. *Mol Cell Biol* 22, 148-160.
- McKeown, M., Belote, J. M., and Baker, B. S. (1987). A molecular analysis of transformer, a gene in *Drosophila melanogaster* that controls female sexual differentiation. *Cell* 48, 489-499.
- Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* 402, 838-841.

- Mermoud, J. E., Cohen, P. T., and Lamond, A. I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *Embo J* 13, 5679-5688.
- Michael, W. M., Siomi, H., Choi, M., Pinol-Roma, S., Nakielny, S., Liu, Q., and Dreyfuss, G. (1995). Signal sequences that target nuclear import and nuclear export of pre-mRNA-binding proteins. *Cold Spring Harb Symp Quant Biol* 60, 663-668.
- Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S., and Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *Embo J* 18, 4308-4320.
- Misteli, T. (2000). Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. *J Cell Sci* 113 (Pt 11), 1841-1849.
- Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol* 143, 297-307.
- Misteli, T., Caceres, J. F., and Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387, 523-527.
- Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. *Nat Genet* 30, 13-19.
- Monsalve, M., Wu, Z., Adelmant, G., Puigserver, P., Fan, M., and Spiegelman, B. M. (2000). Direct coupling of transcription and mRNA processing through the thermogenic coactivator PGC-1. *Mol Cell* 6, 307-316.
- Murray, H. L., and Jarrell, K. A. (1999). Flipping the switch to an active spliceosome. *Cell* 96, 599-602.
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 348, 515-520.
- Nakielny, S., and Dreyfuss, G. (1996). The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *J Cell Biol* 134, 1365-1373.
- Nelissen, R. L., Will, C. L., van Venrooij, W. J., and Luhrmann, R. (1994). The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *Embo J* 13, 4113-4125.
- Nestel, F. P., Colwill, K., Harper, S., Pawson, T., and Anderson, S. K. (1996). RS cyclophilins: identification of an NK-TR1-related cyclophilin. *Gene* 180, 151-155.
- Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., and Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc Natl Acad Sci U S A* 94, 385-390.

- Nilsen, T. W. (1995). trans-splicing: an update. *Mol Biochem Parasitol* 73, 1-6.
- Nilsen, T. W. (2002). The spliceosome: no assembly required? *Mol Cell* 9, 8-9.
- Nishii, Y., Morishima, M., Kakehi, Y., Umehara, K., Kioka, N., Terano, Y., Amachi, T., and Ueda, K. (2000). CROP/Luc7A, a novel serine/arginine-rich nuclear protein, isolated from cisplatin-resistant cell line. *FEBS Lett* 465, 153-156.
- Niwa, M., and Berget, S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* 5, 2086-2095.
- Nogues, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A. R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J Biol Chem* 277, 43110-43114.
- Ohno, M., and Shimura, Y. (1996). A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes Dev* 10, 997-1007.
- Okamoto, Y., Onogi, H., Honda, R., Yasuda, H., Wakabayashi, T., Nimura, Y., and Hagiwara, M. (1998). cdc2 kinase-mediated phosphorylation of splicing factor SF2/ASF. *Biochem Biophys Res Commun* 249, 872-878.
- Ortlepp, D., Laggerbauer, B., Mullner, S., Achsel, T., Kirschbaum, B., and Luhrmann, R. (1998). The mammalian homologue of Prp16p is overexpressed in a cell line tolerant to Leflunomide, a new immunoregulatory drug effective against rheumatoid arthritis. *Rna* 4, 1007-1018.
- Patterson, B., and Guthrie, C. (1991). A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell* 64, 181-187.
- Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev* 7, 393-406.
- Peng, X., and Mount, S. M. (1995). Genetic enhancement of RNA-processing defects by a dominant mutation in B52, the Drosophila gene for an SR protein splicing factor. *Mol Cell Biol* 15, 6273-6282.
- Pinol-Roma, S., and Dreyfuss, G. (1991). Transcription-dependent and transcription-independent nuclear transport of hnRNP proteins. *Science* 253, 312-314.
- Pinol-Roma, S., and Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* 355, 730-732.
- Proudfoot, N. J., Furger, A., and Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell* 108, 501-512.

- Query, C. C., Bentley, R. C., and Keene, J. D. (1989). A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol Cell Biol* 9, 4872-4881.
- Query, C. C., Strobel, S. A., and Sharp, P. A. (1996). Three recognition events at the branch-site adenine. *Embo J* 15, 1392-1402.
- Rando, O. J., Zhao, K., and Crabtree, G. R. (2000). Searching for a function for nuclear actin. *Trends Cell Biol* 10, 92-97.
- Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12, 1231-1245.
- Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* 3, 2113-2123.
- Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev* 6, 215-220.
- Ring, H. Z., and Lis, J. T. (1994). The SR protein B52/SRp55 is essential for *Drosophila* development. *Mol Cell Biol* 14, 7499-7506.
- Rio, D. C. (1988). Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. *Proc Natl Acad Sci U S A* 85, 2904-2908.
- Romfo, C. M., Maroney, P. A., Wu, S., and Nilsen, T. W. (2001). 3' splice site recognition in nematode trans-splicing involves enhancer-dependent recruitment of U2 snRNP. *Rna* 7, 785-792.
- Rossi, F., Labourier, E., Forne, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C., and Tazi, J. (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* 381, 80-82.
- Roth, M. B., Zahler, A. M., and Stolk, J. A. (1991). A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J Cell Biol* 115, 587-596.
- Ruegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell* 1, 243-253.
- Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J., and Wasserman, S. A. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. *Cell* 87, 1079-1089.
- Sanford, J. R., and Bruzik, J. P. (1999). SR proteins are required for nematode trans-splicing in vitro. *Rna* 5, 918-928.
- Sanford, J. R., Longman, D., and Caceres, J. F. (2003). Multiple roles of the SR protein family in splicing regulation. *Prog Mol Subcell Biol* 31, 33-58.

- Schaal, T. D., and Maniatis, T. (1999a). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol* 19, 261-273.
- Schaal, T. D., and Maniatis, T. (1999b). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol* 19, 1705-1719.
- Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., and Mattaj, I. W. (1989). Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *Embo J* 8, 4163-4170.
- Schmidt-Zachmann, M. S., Dargemont, C., Kuhn, L. C., and Nigg, E. A. (1993). Nuclear export of proteins: the role of nuclear retention. *Cell* 74, 493-504.
- Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., and Zipursky, S. L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671-684.
- Schul, W., van Driel, R., and de Jong, L. (1998). A subset of poly(A) polymerase is concentrated at sites of RNA synthesis and is associated with domains enriched in splicing factors and poly(A) RNA. *Exp Cell Res* 238, 1-12.
- Schutt, C., and Nothiger, R. (2000). Structure, function and evolution of sex-determining systems in Dipteran insects. *Development* 127, 667-677.
- Schwer, B., and Gross, C. H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *Embo J* 17, 2086-2094.
- Schwer, B., and Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* 349, 494-499.
- Screaton, G. R., Caceres, J. F., Mayeda, A., Bell, M. V., Plebanski, M., Jackson, D. G., Bell, J. I., and Krainer, A. R. (1995). Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. *Embo J* 14, 4336-4349.
- Seraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *Embo J* 7, 2533-2538.
- Shepard, J., Reick, M., Olson, S., and Graveley, B. R. (2002). Characterization of U2AF(6), a splicing factor related to U2AF(35). *Mol Cell Biol* 22, 221-230.
- Shin, C., and Manley, J. L. (2002). The SR protein SRp38 represses splicing in M phase cells. *Cell* 111, 407-417.
- Shyu, A. B., and Wilkinson, M. F. (2000). The double lives of shuttling mRNA binding proteins. *Cell* 102, 135-138.

- Siliciano, P. G., and Guthrie, C. (1988). 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev* 2, 1258-1267.
- Simons, F. H., Broers, F. J., Van Venrooij, W. J., and Pruijn, G. J. (1996). Characterization of cis-acting signals for nuclear import and retention of the La (SS-B) autoantigen. *Exp Cell Res* 224, 224-236.
- Smith, C. W., and Nadal-Ginard, B. (1989). Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell* 56, 749-758.
- Smith, C. W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* 25, 381-388.
- Spector, D. L., O'Keefe, R. T., and Jimenez-Garcia, L. F. (1993). Dynamics of transcription and pre-mRNA splicing within the mammalian cell nucleus. *Cold Spring Harb Symp Quant Biol* 58, 799-805.
- Spritz, R. A., Strunk, K., Surowy, C. S., Hoch, S. O., Barton, D. E., and Francke, U. (1987). The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res* 15, 10373-10391.
- Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Staley, J. P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell* 3, 55-64.
- Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell* 9, 31-44.
- Sukegawa, J., and Blobel, G. (1995). A putative mammalian RNA helicase with an arginine-serine-rich domain colocalizes with a splicing factor. *J Biol Chem* 270, 15702-15706.
- Sun, C. T., Lo, W. Y., Wang, I. H., Lo, Y. H., Shiou, S. R., Lai, C. K., and Ting, L. P. (2001). Transcription repression of human hepatitis B virus genes by negative regulatory element-binding protein/SON. *J Biol Chem* 276, 24059-24067.
- Surowy, C. S., van Santen, V. L., Scheib-Wixted, S. M., and Spritz, R. A. (1989). Direct, sequence-specific binding of the human U1-70K ribonucleoprotein antigen protein to loop I of U1 small nuclear RNA. *Mol Cell Biol* 9, 4179-4186.
- Sutherland, H. G., Mumford, G. K., Newton, K., Ford, L. V., Farrall, R., Dellaire, G., Caceres, J. F., and Bickmore, W. A. (2001). Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Hum Mol Genet* 10, 1995-2011.

- Tacke, R., Chen, Y., and Manley, J. L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc Natl Acad Sci U S A* 94, 1148-1153.
- Tacke, R., and Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *Embo J* 14, 3540-3551.
- Tacke, R., and Manley, J. L. (1999). Determinants of SR protein specificity. *Curr Opin Cell Biol* 11, 358-362.
- Tanner, S., Stagljar, I., Georgiev, O., Schaffner, W., and Bourquin, J. P. (1997). A novel SR-related protein specifically interacts with the carboxy-terminal domain (CTD) of RNA polymerase II through a conserved interaction domain. *Biol Chem* 378, 565-571.
- Tarn, W. Y., and Steitz, J. A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev* 8, 2704-2717.
- Tarn, W. Y., and Steitz, J. A. (1995). Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. *Proc Natl Acad Sci U S A* 92, 2504-2508.
- Tate, P., Lee, M., Tweedie, S., Skarnes, W. C., and Bickmore, W. A. (1998). Capturing novel mouse genes encoding chromosomal and other nuclear proteins. *J Cell Sci* 111 (Pt 17), 2575-2585.
- Teigelkamp, S., Mundt, C., Achsel, T., Will, C. L., and Luhrmann, R. (1997). The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *Rna* 3, 1313-1326.
- Teigelkamp, S., Whittaker, E., and Beggs, J. D. (1995). Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. *Nucleic Acids Res* 23, 320-326.
- Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Luhrmann, R., and Philipson, L. (1986). Cloning of the human cDNA for the U1 RNA-associated 70K protein. *Embo J* 5, 3209-3217.
- Tronchere, H., Wang, J., and Fu, X. D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* 388, 397-400.
- Umen, J. G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. *Rna* 1, 869-885.
- Vagner, S., Vagner, C., and Mattaj, I. W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev* 14, 403-413.
- Valadkhan, S., and Manley, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* 413, 701-707.

- Valcarcel, J., Gaur, R. K., Singh, R., and Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* 273, 1706-1709.
- van der Houven van Oordt, W., Diaz-Meco, M. T., Lozano, J., Krainer, A. R., Moscat, J., and Caceres, J. F. (2000). The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J Cell Biol* 149, 307-316.
- van Der Houven Van Oordt, W., Newton, K., Screatton, G. R., and Caceres, J. F. (2000). Role of SR protein modular domains in alternative splicing specificity in vivo. *Nucleic Acids Res* 28, 4822-4831.
- Vijayraghavan, U., and Abelson, J. (1990). PRP18, a protein required for the second reaction in pre-mRNA splicing. *Mol Cell Biol* 10, 324-332.
- Vijayraghavan, U., Company, M., and Abelson, J. (1989). Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*. *Genes Dev* 3, 1206-1216.
- Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol* 140, 737-750.
- Wang, H. Y., Xu, X., Ding, J. H., Bermingham, J. R., Jr., and Fu, X. D. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. *Mol Cell* 7, 331-342.
- Wang, J., and Manley, J. L. (1995). Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways. *Rna* 1, 335-346.
- Wang, J., Takagaki, Y., and Manley, J. L. (1996). Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev* 10, 2588-2599.
- Wang, P., Lou, P. J., Leu, S., and Ouyang, P. (2002). Modulation of alternative pre-mRNA splicing in vivo by pinin. *Biochem Biophys Res Commun* 294, 448-455.
- Whittaker, E., and Beggs, J. D. (1991). The yeast PRP8 protein interacts directly with pre-mRNA. *Nucleic Acids Res* 19, 5483-5489.
- Will, C. L., and Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol* 9, 320-328.
- Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061-1070.
- Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402, 832-835.

- Wynn, S. L., Fisher, R. A., Pagel, C., Price, M., Liu, Q. Y., Khan, I. M., Zammit, P., Dadrah, K., Mazrani, W., Kessling, A., *et al.* (2000). Organization and conservation of the GART/SON/DONSON locus in mouse and human genomes. *Genomics* 68, 57-62.
- Xiao, S. H., and Manley, J. L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 11, 334-344.
- Xiao, S. H., and Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *Embo J* 17, 6359-6367.
- Yun, C. Y., Velazquez-Dones, A. L., Lyman, S. K., and Fu, X. D. (2003). Phosphorylation-dependent and -independent nuclear import of RS domain-containing splicing factors and regulators. *J Biol Chem* 278, 18050-18055.
- Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R. V., Gentile, C., Gebara, M., and Corden, J. L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A* 93, 6975-6980.
- Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* 6, 837-847.
- Zahler, A. M., Neugebauer, K. M., Stolk, J. A., and Roth, M. B. (1993). Human SR proteins and isolation of a cDNA encoding SRp75. *Mol Cell Biol* 13, 4023-4028.
- Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg, S. L., Person, C., and Sands, A. T. (1998). Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* 392, 608-611.
- Zamore, P. D., and Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* 86, 9243-9247.
- Zamore, P. D., Patton, J. G., and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355, 609-614.
- Zeng, C., Kim, E., Warren, S. L., and Berget, S. M. (1997). Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *Embo J* 16, 1401-1412.
- Zerler, B., Moran, B., Maruyama, K., Moomaw, J., Grodzicker, T., and Ruley, H. E. (1986). Adenovirus E1A coding sequences that enable ras and pmt oncogenes to transform cultured primary cells. *Mol Cell Biol* 6, 887-899.
- Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I., and Green, M. R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A* 89, 8769-8773.

- Zhang, W. J., and Wu, J. Y. (1996). Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Mol Cell Biol* 16, 5400-5408.
- Zhang, W. J., and Wu, J. Y. (1998). Sip1, a novel RS domain-containing protein essential for pre-mRNA splicing. *Mol Cell Biol* 18, 676-684.
- Zheng, Z. M., Huynen, M., and Baker, C. C. (1998). A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly. *Proc Natl Acad Sci U S A* 95, 14088-14093.
- Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182-185.
- Zhou, Z., and Reed, R. (1998). Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *Embo J* 17, 2095-2106.
- Zhu, J., Mayeda, A., and Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8, 1351-1361.
- Zhuang, Y., and Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46, 827-835.
- Zorio, D. A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 402, 835-838.
- Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev* 10, 1356-1368.